

[illegible]

LOCUS	DEFINITION	925 bp	DNA	linear	GSS 03-JUN-1999
CNS0091P/c					
CNS0091P	Drosophila melanogaster genome survey sequence TE13 end of BAC # BACR15D16 of Rpcl-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				

ACCESSION	AL053013	GI:4934461
VERSION	AL053013.1	
KEYWORDS	GSS.	
SOURCE	fruit fly.	
ORGANISM	<i>Drosophila melanogaster</i>	

REFERENCE
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 925)
Genoscope,
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage ::

COMMENT

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammosser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPc1-98 and was constructed by partial EcoRI digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACRPC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	Location/Qualifiers
source	1. .925

BASE COUNT	120 a	61 c	61 g	172 t	511 others
ORIGIN					
Query Match		12.5%	Score 49.6;	DB 12;	Length 925;
Best Local Similarity		12.8%	Pred. NO. 0.75;		
Matches 43;	Conservative 157;	Mismatches 136;	Indels 0;	Gaps 0;	

[illegible]

Qy	Dy
300	720
350	770
400	820
450	870
500	920
550	970
600	1020
650	1070
700	1120
750	1170
800	1220
850	1270
900	1320
950	1370
1000	1420
1050	1470
1100	1520
1150	1570
1200	1620
1250	1670
1300	1720
1350	1770
1400	1820
1450	1870
1500	1920
1550	1970
1600	2020
1650	2070
1700	2120
1750	2170
1800	2220
1850	2270
1900	2320
1950	2370
2000	2420
2050	2470
2100	2520
2150	2570
2200	2620
2250	2670
2300	2720
2350	2770
2400	2820
2450	2870
2500	2920
2550	2970
2600	3020
2650	3070
2700	3120
2750	3170
2800	3220
2850	3270
2900	3320
2950	3370
3000	3420
3050	3470
3100	3520
3150	3570
3200	3620
3250	3670
3300	3720
3350	3770
3400	3820
3450	3870
3500	3920
3550	3970
3600	4020
3650	4070
3700	4120
3750	4170
3800	4220
3850	4270
3900	4320
3950	4370
4000	4420
4050	4470
4100	4520
4150	4570
4200	4620
4250	4670
4300	4720
4350	4770
4400	4820
4450	4870
4500	4920
4550	4970
4600	5020
4650	5070
4700	5120
4750	5170
4800	5220
4850	5270
4900	5320
4950	5370
5000	5420
5050	5470
5100	5520
5150	5570
5200	5620
5250	5670
5300	5720
5350	5770
5400	5820
5450	5870
5500	5920
5550	5970
5600	6020
5650	6070
5700	6120
5750	6170
5800	6220
5850	6270
5900	6320
5950	6370
6000	6420
6050	6470
6100	6520
6150	6570
6200	6620
6250	6670
6300	6720
6350	6770
6400	6820
6450	6870
6500	6920
6550	6970
6600	7020
6650	7070
6700	7120
6750	7170
6800	7220
6850	7270
6900	7320
6950	7370
7000	7420
7050	7470
7100	7520
7150	7570
7200	7620
7250	7670
7300	7720
7350	7770
7400	7820
7450	7870
7500	7920
7550	7970
7600	8020
7650	8070
7700	8120
7750	8170
7800	8220

RESULT	5
BH020987	
LOCUS	
DEFINITION	
BH020987	561 bp DNA linear GSS-25-MAY-2001
P864c.d.LEISHAPAC-left.1	Leishmania major Friedlin Cosmid Genomic
Library Leishmania major genomic clone P864c.	DNA sequence.

ACCESSION	BH020987
VERSION	BH020987.1
KEYWORDS	GI:14202102
SOURCE	GSS.
ORGANISM	leishmania major.
	leishmania major

REFERENCE	AUTHORS	TITLE
1 (bases 1 to 561)		
Myler, P. J., Vogt, C.,	Cawthra, J., Klacking, M., Marty, A., Mack, J.,	
Munden, H., Nguyen, D.,	Robertson, L., Sik, E., Paellinla, G., Aggarwal,	
G., Nelson, S., Seyler, A.,	Worthy, E., and Stuart, K.	
Leishmania major	Friedlin	Cosmid End Sequences

COMMENT

Contact: Myler PJ
Seattle Biomedical Research Institute
4 Nickerson Street, Seattle, WA 98109-1651, USA
Tel: 206 284-8846
Fax: 206 284-0313
Email: mylerpj@sbri.org
Seq primer: LEISHPAC-left
Class: PAC end.

FEATURES	Location/Qualifiers
source	1. .561

```

/organism="Leishmania major"
/strain="Friedlin"
/db_xref="taxon:5664"
/clone="P864c"
/clone_11b="Leishmania major Friedlin Cosmid Genomic
library"
/lab_host="E. coli ED8767"
/note="Vector: cLHYG; Site:1: BamHI: Genomic DNA from
Leishmania major Friedlin was partially digested with
Sma31, size selected, and ligated with BamHI-digested
cLHYG cosmid vector DNA. 9216 clones were picked and
arrayed. Library construction is described in Ivens et
al., Genomics Research, 8:135-145 (1998). The cLHYG
vector (Acc. No. CV059231) is described in Ryan et al.,
Gene, 131:145-150 (1993)"

```

Query Match	12.4%;	Score 49.2;	DB 12;	Length 561;
Best Local Similarity	63.6%;	Pred. No. 0.9;		
Matches	75;	Conservative	0;	Mismatches 43; Indels 0; Gaps 0;
Oy	159	caacaacgcaacgcgcgacgaagtcacacgcgttgctcggaagcgctcgcggcgcaagtct	218	
Db	13	CGAGCATGGGCGATCGCCGGGGCCCTGGTATCTCAAGGCGCCGTCGGGATCCCGGCTCCGG	72	
Oy	219	cggatcttcacacggcggaactgtacacggcgagtcgaagcgcgcttcgaataactcgagcc	276	
Db	73	CGGATTCACACCCGGGACGCTATCTCGCCCTTCAACGCGACGCGCATCAAGTCAAGTAC	130	
RESULT	6			

	BE361689/c	634 bp	mRNA	linear	EST 20-JUL-2000
LOCUS	DGJ_81.E06.g1_A002 Dark Grown 1 (DGJ)	Sorghum bicolor cDNA,	mRNA		
DEFINITION	sequence.				
ACCESSION	BE361689				
VERSION	BE361689.1	GI:9303246			
KEYWORDS	EST.				
SOURCE	sorghum.				
ORGANISM	Sorghum bicolor				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoidae; Andropogoneae; Sorghum. 1 (bases 1 to 634) Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt, L.H.				
AUTHORS	An EST database from Sorghum: dark-grown seedlings Unpublished (2000) Contact: Cordonnier-Pratt MM Department of Botany The University of Georgia Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel.: 706 542 1860 Fax: 706 542 1805 Email: mmp@atluga.edu				
TITLE	Sequences have been trimmed to exclude polyA, vector and regions below paired quality 16. The threshold for highest quality sequence is 20.				
JOURNAL	Seq primer: polyTMix				
COMMENT	High quality sequence start: 67 High quality sequence stop: 634 POLYA=NO.				
	Location/Qualifiers				
FEATURES	1..634				
Source	/organism="Sorghum bicolor" /db_xref="taxon:4558"				
	/clone_lib="Dark Grown 1 (DGJ)" /note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda zap; Site.1: XhoI; Site.2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda Zap II. Clones to be sequenced were prepared by mass excision."				
BASE COUNT	122 a 198 c 212 g 102 t				
ORIGIN					

	Query Match:	12.3%	Score	48.6;	DB	10;	Length	634;	
	Best Local Similarity:	47.0%;	Pred.	No. 1.2;	Mismatches	169;	Indels	0;	Gaps
	Matches	150;	Conservative	0;					
Dy	17	acttcacgctgtccaccagggtlgygcagaygatltcgcatcgcattcgatcggycagagcgaatlgcga	76						
Dd	388	ACGAGCGGCCGTCCACGAATGATGGCAACGCCTTGTCGGCCGCGAGGCCCTGGCCCGTGGCGA	329						
OY	77	tgcgcggccaagatccgatcgttgtvggysgttaaccaacctcaatatcggactaacgctt	136						
Dd	328	ACCCTTTCCCGCAGATGTGCACCCGCTGGGCCCAACGGTAGAAATCTCCTCCACG	269						
OY	137	ttcttcgagttygttgatgcacaanaacgaaagcggcacagatccaacgcgtgtgtcg	156						
Dd	268	ACGACGTGTCGTGTCGTGTCGTCTCTCGGAATCGGGCGGAAAACCTCTCTGCTGCT	209						
OY	197	ggagagcttcgcggcgaafctcgcgcatctccaccgscggagactgataccaacgcgtlcagc	256						
Dd	208	GCGCGCGCGGGGGGGGGGTACTGCTCCGGGTGGGGCTTGCGGTGCGGTGACATTGT	149						
OY	257	gcgcctccgatcaaactcgcccaccgcgaatgcyggacgcgcttaaagcgcatatcccggtg	316						
Dd	148	GCGCGCGCAACGCGCTGTGTAACAAGCGCAAGCCCTTCGCCAACACCGCGCACCGGAACCGGA	89						
OY	317	acgtcatcctgctgfaacctg	335						
Dd	*88	GCTGCTCTGTGTGGCGCAG	70						

FEATURES	LOCATION/Qualifiers
RESULT	289 bp DNA linear GSS 25-JUL-2000
P664L	
LOCUS	
DEFINITION	Leishmania major Friedlin PAC P664 left end-sequence, genomic survey sequence.
ACCESSION	AL390548
VERSION	AL390548.1 GI:9501524
KEYWORDS	GSS.
SOURCE	Leishmania major.
ORGANISM	Leishmania major Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
REFERENCE	Leishmania
AUTHORS	1 (bases 1 to 289) Ivens,A.C., Lewis,S.M., Bagherzadeh,A., Zhang,L., Chan,H.M. and Smith,D.F.
TITLE	A physical map of the Leishmania major Friedlin genome
JOURNAL	Genome Res. 8 (2), 135-145 (1998)
MEDLINE	98146435
REFERENCE	2 (bases 1 to 289)
AUTHORS	Taylor,R.G., Huckle,E.E.J., Ivens,A.C., Rajandream,M.A. and Barrell,B.G.
TITLE	Direct Submission
JOURNAL	Submitted (24-JUL-2000) Leishmania major Friedlin genome sequencing project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and allicates@sanger.ac.uk
COMMENT	see http://www.ebi.ac.uk/parasites/leish.html Details of Leishmania sequencing at the Sanger Centre are available at http://www.sanger.ac.uk/Projects/Lmajor/ The primer sequence can be obtained from allicates@sanger.ac.uk.

```

SOURCE
    I. .209
    /organism="Leishmania major"
    /strain="Friedlin"
    /db_xref="taxon:5664"
    /clone="PAC p864"

BASE COUNT      49 a      86 c      108 g      46 t

ORIGIN

Query Match      12.0% ; Score 47.6 ; DB 12 ; Length 289 ;
Best Local Similarity 62.7% ; Pred. No. 1.8 ;
Matches 74 ; Conservative 0 ; Mismatches 44 ; Indels 0 ; Gaps 0 ;

Qy      159 caataaaggaagagcgacagatccaacgagtgtagcgagcgctccggcggcaagtct 218
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      24 CGAAGCGGATCGCGCGCGGCGCCTGTATATGAGAGGCCGCTGGGCTCCCGGCTCCGT 83
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Qy      219 cggatctccaccgcgcgacglatcacccgagtcgagcgagcgtccgaactcagcc 276
          ||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      84 CGGATTCACACCGCGGAGCGTGTGTCGTGACCGGCTTCACACGCGCATCAAGTCAGTC 141
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT      8
CONS006XK
LOCUS
DEFINITION
    CNS006XK          935 bp      DNA          linear      GSS 03-JUN-1999
VERSION
    BACR41N09 of RPCI-98 library from Drosophila melanogaster (fruit
KEYWORDS
    fly), genomic survey sequence.
ACCESSION
    AL066051
VERSION
    AL066051.1      GI:4945019
KEYWORDS
    GSS.
SOURCE
    fruit fly.
ORGANISM
    Drosophila melanogaster
    Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
    Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
    Muscomorpha; Ephydriidae; Drosophilidae; Drosophila.
    1 (bases 1 to 935)
REFERENCE
    Genoscope.
    Direct Submission
    Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : :
    BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
    - Web : www.genoscope.cns.fr)

```

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Oseguera and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2: cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

1. .935
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPI-98"
/clone="BAC1409"
/note="end : 17"

BASE COUNT 257 a 170 c 162 g 96 t 250 others
ORIGIN

Query Match 12.0% Score 47.4; DB 12; Length 935;
Best Local Similarity 28.2% Pred. No. 2.3; Mismatches 193; Indels 0; Gaps 0;

Matches 111; Conservative 89; Mismatches 193; Indels 0; Gaps 0;

QY 4 gccgcgtccgataactccagctccagaggtgagcaggttcgcatccgacgag 63
DB 523 GCSTCCMYMSSVSCSCGTCGCGTSCMCMCTGCKGCGGCTGSCSSSSCS 582
QY 64 caggcagatgcagtcgagcagcagcagcagcagcagcagcagcagcagc 123
DB 583 BBVSTVCSCTBCKKSSGCTGSCGCGGCGGCGGCGGCGGCGGCGGCGG 642
QY 124 gggcctacgcgctctcctcgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 183
DB 643 GSSSGSSGSSSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSGSSG 702
QY 184 caacgcgtgtcgtgagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 243
DB 703 GSGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 762
QY 244 accgcgtgtcagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 303
DB 763 MSMSGCGGSSGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGSS 822
QY 304 catcatccggtgagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtc 363
DB 823 GSGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 882
QY 364 gggagcgtgacattgagcagcagcagcagcagcagcagcagcagcagc 396
DB 883 GSGSGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 915

RESULT 9

CNS010EM/c 1009 bp DNA linear GSS 26-JUL-1999
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
BACN03p19 of Drosophila library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL098882
VERSION AL098882.1 GI:5610493
KEYWORDS
SOURCE
ORGANISM

fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

TITLE

JOURNAL

1 (bases 1 to 1009)
Genoscope.
Direct Submission
Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
<http://www.edgp.ebi.ac.uk> - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billaud at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelBAC11.

FEATURES

source

1. .1009
/organism="Drosophila melanogaster"
/plasmid="pBelBAC11"
/db_xref="taxon:7227"
/clone_lib="DrosBAC"
/clone="BACN03p19"
/note="end : SP6"

BASE COUNT 147 a 377 c 178 g 64 t 243 others
ORIGIN

Query Match 11.9% Score 47; DB 12; Length 1009;
Best Local Similarity 30.8% Pred. No. 2.8;

Matches 111; Conservative 80; Mismatches 168; Indels 1; Gaps 1;

QY 37 gggcagagatccgcatccgacgagcagcagcagcagcagcagcagcagc 96
DB 1002 RSGSVSSGMAVVCVCGCGSARCGAASCACGCMACACGCGCMRAGSAATCGGAGCC 943
QY 97 ggtggggtgacccacccagcttaccatcagcagcagcagcagcagcagc 156
DB 942 GCGCMGSGAGCRSCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 883
QY 157 gacacacagcagcagcagcagcagcagcagcagcagcagcagcagcag 216
DB 882 RSMGCGGCGGSSGSGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 823
QY 217 ctgcgcatccacacgcgagcagcagcagcagcagcagcagcagcagcagc 276
DB 822 SSSCGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 763
QY 277 accgcgtgtcagcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgtcgt 336
DB 762 GSSGCGGCGGSSGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 704
QY 337 caaacagtcgagcagcagcagcagcagcagcagcagcagcagcagcagc 396
DB 703 CSMKCSVCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 644

RESULT 10

BE215641 538 bp mRNA linear EST 23-OCT-2001
LOCUS
DEFINITION HV.CEP0007J02f Hordeum vulgare seedling green leaf EST library
HV.CEP0007J02f, mRNA sequence.

ACCESSION BE215641
VERSION BE215641.1 GI:8903169
KEYWORDS
SOURCE
ORGANISM

Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
1 (bases 1 to 538)
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Wei, F., Begum, D.,
Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, D., Choi,
D.W., Fenton, R.D., Oates, R. and Main, D.

Tue Aug 6 11:59:24 2002

us-09-684-215a-3.rst

Page 9

QY 301 gggcaccatccgggtgacgtatcctggtgacatggaacccaatcgaggcgacgct 360
|||||
Db 152284 gggcaccatccgggtgacgtatcctggtgacatggaacccaatcgaggcgacgct 152343
QY 361 acaggaacgtgacatggtgacgaggaaccccgccg 396
|||||
Db 152344 acaggaacgtgacatggtgacgaggaaccccgccg 152379

RESULT 2

US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 100.0%; Score 396; DB 4; Length 4411529;
Best Local Similarity 100.0%; Pred. No. 3.1e-91;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acggcgcgctccgataactcagctgctccaggtgagggagagatcgcatccgac 60
|||||
Db 151815 accgcgcgctccgataactcagctgctccaggtgagggagagatcgcatccgac 151874
QY 61 gggcagcgatgagcgatcgagggcgagatccgagtggtgggtggtaccacggtcat 120
|||||
Db 151875 gggcagcgatgagcgatcgagggcgagatccgagtggtgggtggtaccacggtcat 151934
QY 121 atcgggcgctacggccttcctcgctggtggtgtgtcgcacaacagcgcaacgagcgca 180
|||||
Db 151935 atcgggcgctacggccttcctcgctggtggtgtgtcgcacaacagcgcaacgagcgca 151994
QY 181 gttcaacggtggtggtgagcggtcgcggcgcaagtcctcgcatccacggcgagctg 240
|||||
Db 151995 gttcaacggtggtggtgagcggtcgcggcgcaagtcctcgcatccacggcgagctg 152054
QY 241 atcacggcggtcagcgcggtcctcgatcaactcgccacggcgagtgcgagcgcttaac 300
|||||
Db 152055 atcacggcggtcagcgcggtcctcgatcaactcgccacggcgagtgcgagcgcttaac 152114
QY 301 gggcaccatccgggtgacgtatcctggtgacatggaacccaatcgaggcgacgct 360
|||||
Db 152115 gggcaccatccgggtgacgtatcctggtgacatggaacccaatcgaggcgacgct 152174
QY 361 acaggaacgtgacatggtgacgaggaaccccgccg 396
|||||
Db 152175 acaggaacgtgacatggtgacgaggaaccccgccg 152210

RESULT 3

US-08-818-112-4
; Sequence 4, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Mark, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-4

Query Match 99.6%; Score 394.4; DB 4; Length 447;
Best Local Similarity 99.7%; Pred. No. 1.5e-91;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggcgcgctccgataactcagctgctccaggtgagggagagatcgcatccgac 60
|||||
Db 11 ACAGCGCGCTCCGATAACTCAGCTGCTCCAGGTTGGCGAGGATTCGCCATTCCGATC 70
QY 61 gggcagcgatgagcgatcgagggcgagatccgagtggtgggtggtaccacggtcat 120
|||||
Db 71 GGGCAGCGATGAGCGATCGGCGCGAGTCGATCGAGTGGGGGTACCCACCGCTTAT 130
QY 71 gggcagcgatgagcgatcgagggcgagatccgagtggtgggtggtaccacggtcat 120
|||||
Db 71 GGGCAGCGATGAGCGATCGGCGCGAGTCGATCGAGTGGGGGTACCCACCGCTTAT 130
QY 121 atcgggcgctacggccttcctcgctggtggtgtgtcgcacaacagcgcaacgagcgca 180
|||||
Db 131 ATCGGGCGCTACCGCTTCCTCGGCTTGGGTGTGTGCAACAACGCGACGCGACGA 190
QY 181 gttcaacggtggtggtgagcggtcgcggcgcaagtcctcgcatccacggcgagctg 240
|||||
Db 191 GttCAAGGCTGgtGgtGgAGCGctCCGCGCAAGTCTGCGCATCTCCACGGGAGCTG 250
QY 241 atcacggcggtcagcgcggtcctcgatcaactcgccacggcgagtgcgagcgcttaac 300
|||||
Db 251 ATCACGGCGGTGACGCGGCTCCGATCAACTCGGCGCACCGCGATGGGCGAGCGCTTAA 310
QY 301 gggcaccatccgggtgacgtatcctggtgacatggaacccaatcgaggcgacgct 360
|||||
Db 311 GGGCATCTATCCGATGAGCGTATCTCGTGAACCTGGCAACCAAGTGGGCGGACGCGT 370
QY 361 acaggaacgtgacatggtgacgaggaaccccgccg 396
|||||
Db 371 ACAGGAACGTGACATGgtGCGCGAGGAGACCCCGGCG 406

RESULT 4
US-08-818-111-4
; Sequence 4, Application US/08818111
; Patent No. 633852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818.111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 682-6031
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-111-4

Query Match 99.6%; Score 394.4; DB 4; Length 447;
Best Local Similarity 99.7%; Pred. No. 1.5e-91;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 acggcgcgctccgataactccagctgtcccaagggtggcgaagattcgccattccgac 60
DB 11 ACGGCGCGCTCCGATAACTCCAGCTGTCCCAAGGTTGGCGAGGATTCGCCATTCCGATC 70
QY 61 gggcgaagcagatgagctgcggcgagcagatccgagtggtgggggttcaaccaccgttcat 120
DB 71 GGGCAGGCGATGCGATGCGGCGGCGAGATCGATCGAGTGGGGGTCAACCCACCGTTTCA 130
QY 121 atcgagcctacgcgctccctccgcttggtgtgtgtcgacacaaagcgaagcgacga 180
DB 131 ATCGGCGCTACGCCCTTCTCGGCTTGAGTGTTCGACAAACGCGACACGCGGCGACGA 190
QY 181 gtccaacgctgtgtcggagagcgttcggcgagcgaagtcgtgcatctccaccgsgagctg 240
DB 191 GTCCAACGCTGTGTGGGAGCGCTCCGCGGCGCAAGTCTCGCATCTCACCGGCGAGCTG 250
QY 241 atcaacgcgctcgaagcgctcccgatcaactcggccacgcgagtgagcgagcgcttaac 300
DB 251 ATCAACGCGCTCGACGCGCTCCGATCAACTCGGCGCACCGGAGTGGGAGCGCTTAAC 310
QY 301 gggcatcatcccggtgacgtcatctcggtgacactgcaaaccaagtcggcgagcgct 360
|||||

DB 311 GGGCATCATCCCGGTGACGTCATCTCGTGAATGCGCAACCAAGTCGGGGCGACGGCT 370
QY 361 acagggaacgtgacatggtccgagagagaccccgcc 396
DB 371 ACAGGAGACGTGACATTGGCCGAGGAGACCCCGGCC 406
RESULT 5
US-09-056-556-4
; Sequence 4, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056.556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-4

Query Match 99.6%; Score 394.4; DB 4; Length 447;
Best Local Similarity 99.7%; Pred. No. 1.5e-91;
Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 acggcgcgctccgataactccagctgtcccaagggtggcgaagattcgccattccgac 60
DB 11 ACGGCGCGCTCCGATAACTCCAGCTGTCCCAAGGTTGGCGAGGATTCGCCATTCCGATC 70
QY 61 gggcgaagcagatgagctgcggcgagcagatccgagtggtgggggttcaaccaccgttcat 120
DB 71 GGGCAGGCGATGCGATGCGGCGGCGAGATCGATCGAGTGGGGGTCAACCCACCGTTTCA 130
QY 121 atcgagcctacgcgctccctccgcttggtgtgtgtcgacacaaagcgaagcgacga 180
DB 131 ATCGGCGCTACGCCCTTCTCGGCTTGAGTGTTCGACAAACGCGACACGCGGCGACGA 190
QY 181 gtccaacgctgtgtcggagagcgttcggcgagcgaagtcgtgcatctccaccgsgagctg 240
DB 191 GTCCAACGCTGTGTGGGAGCGCTCCGCGGCGCAAGTCTCGCATCTCACCGGCGAGCTG 250
QY 241 atcaacgcgctcgaagcgctcccgatcaactcggccacgcgagtgagcgagcgcttaac 300
DB 251 ATCAACGCGCTCGACGCGCTCCGATCAACTCGGCGCACCGGAGTGGGAGCGCTTAAC 310
|||||

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 6, 2002, 08:56:55 ; Search time 1841.62 Seconds
(without alignments)
4499.793 Million cell updates/sec

Title: US-09-684-215A-3

Sequence: 1 acgcccgcgtccgataact.....tgcgcgagggagccccgcgc 396

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

GenBank:*

1: gb_ba:*

2: gb_hg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_hg_hum:*

31: em_hg_inv:*

32: em_hg_other:*

33: em_hggo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Match Length	DB	ID	Description
------------	-------	--------------	----	----	-------------

1	396	100.0	1068	6	AX005788	AX005788 Sequence
2	396	100.0	1143	6	AX005790	AX005790 Sequence
3	396	100.0	11700	1	MFC1418B	296071 Mycobacteri
4	396	100.0	14029	1	AE006925	AE006925 Mycobacte
5	394.4	99.6	447	6	AR169152	AR169152 Sequence
6	394.4	99.6	447	6	BD006325	BD006325 Compounds
7	394.4	99.6	447	6	BD006445	BD006445 Compounds
8	391.2	98.8	1872	6	AR169165	AR169165 Sequence
9	391.2	98.8	1872	6	BD006338	BD006338 Compounds
10	391.2	98.8	1872	6	BD006458	BD006458 Compounds
11	354	89.4	675	6	AX201049	AX201049 Sequence
12	354	89.4	675	6	AX267848	AX267848 Sequence
13	354	89.4	675	6	AX351489	AX351489 Sequence
14	354	89.4	915	6	AX201061	AX201061 Sequence
15	354	89.4	915	6	AX267860	AX267860 Sequence
16	354	89.4	1203	6	AX201078	AX201078 Sequence
17	354	89.4	1203	6	AX267877	AX267877 Sequence
18	354	89.4	1464	6	AX156105	AX156105 Sequence
19	354	89.4	1557	6	AX156089	AX156089 Sequence
20	354	89.4	1578	6	AX155945	AX155945 Sequence
21	354	89.4	1590	6	AX316986	AX316986 Sequence
22	354	89.4	1752	6	AX156109	AX156109 Sequence
23	354	89.4	1752	6	AX156093	AX156093 Sequence
24	354	89.4	1860	6	AX156065	AX156065 Sequence
25	354	89.4	1896	6	AX156081	AX156081 Sequence
26	354	89.4	1941	6	AX156073	AX156073 Sequence
27	354	89.4	1965	6	AX156097	AX156097 Sequence
28	354	89.4	2052	6	AX156113	AX156113 Sequence
29	354	89.4	2076	6	AX156069	AX156069 Sequence
30	354	89.4	2103	6	AX156101	AX156101 Sequence
31	354	89.4	2148	6	AX156077	AX156077 Sequence
32	354	89.4	2148	6	AX156085	AX156085 Sequence
33	354	89.4	2148	6	AX156085	AX156085 Sequence
34	199.2	50.3	1110	1	MP34KDPAG	Z23092 M.paratuber
35	187	47.2	37040	1	MSGB26CS	L7816 Mycobacteri
36	68.6	17.3	37304	1	MLEPRN10	AL583926 Mycobacte
37	68.6	17.3	344050	1	MLEPRN10	AL583926 Mycobacte
38	62.2	15.7	134426	14	1H1CG	AL583917 Mycobacte
39	59	14.9	1395	6	AX005764	M75136 Ictialurid h
40	59	14.9	1771	6	AR169161	AX005764 Sequence
41	59	14.9	1771	6	BD006334	AR169161 Sequence
42	59	14.9	1771	6	BD006454	BD006334 Compounds
43	59	14.9	1812	6	AX005766	BD006454 Compounds
44	59	14.9	16150	1	MTV044	AX005766 Sequence
45	59	14.9	16168	1	AE006985	AL021999 Mycobacte

ALIGNMENTS

RESULT 1

AX005788 1068 bp DNA linear PAT 24-AUG-2000

LOCUS Sequence 907 from Patent WO9090186.

DEFINITION AX005788

ACCESSION AX005788

VERSION AX005788.1 GI:9928795

KEYWORDS

SOURCE

ORGANISM

Mycobacterium tuberculosis.

Mycobacterium tuberculosis

Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;

Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;

Mycobacterium; Mycobacterium tuberculosis complex.

1 (bases 1 to 1068)

Portnoi, D. and Guigueno, A.

Polyptide nucleic sequences exported from mycobacteria, vectors

comprising same and uses for diagnosing and preventing tuberculosis

Patent: WO 9090186-A 907 25-FEB-1999;

PORNOI DENIS (FR); GUIGUENO AGNES (FR)

Location/Qualifiers

FEATURES

source

1..1068

/organism="Mycobacterium tuberculosis"

/db_xref="taxon:1773"

1..1068

CDS

```
/note="50D"  
/codon_start=1  
/transl_table=11  
/protein_id="CAC05178.1"  
/db_xref="GI:9928796"  
/translation="MSNSRRSLRMSWLLSVLAAGLGLATAPQAAPALSGDREAD  
FALPLDPSAMVAOVPQVAVNTKLGNNNAAGAGIYIDNGVLTNNHVLGATID  
INAFVSGQYVGVVGVVADVAIQLRGAGLPSAIGGVAVGEPVAMNGS  
GGGTPRAVPGRVVAVGOTVQASDLSLTGAEETLNLGIOPDAIOPDSDGPPVNGIG  
VGMNTAASDNFOLSGGGGFAPIGQAMAIAGQIRSSGGSPVHIGPTAFGLGIVD  
NNGNARVORVVASPAASLIGTSDGVITAVGAPRINSATMAADLNHHPDVISVT  
WQTKSGRTGTAVTAEPRPA"
```

BASE COUNT 169 a 352 c 375 g 172 t
ORIGIN

Query Match 100.0%; Score 396; DB 6; Length 1068;
Best Local Similarity 100.0%; Pred. No. 7.4e-56;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 acggccgctgcgataacttcagctgctccaggtgaggagagatcgccattccgac 60  
DB 670 ACggccgctgcgataacttcagctgctccaggtgaggagagatcgccattccgac 729  
QY 62 gggcagcgatgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120  
DB 730 Gggcagcgatgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 789  
QY 121 atcgagcctacgccttcctcgcgtggtgtgtgtgcacaaacagcgagcgacga 180  
DB 790 Atcgagcctacgccttcctcgcgtggtgtgtgtgcacaaacagcgagcgacga 849  
QY 181 gtccaaacgctgctgagcgagcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
DB 850 Gtccaaacgctgctgagcgagcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 909  
QY 241 ataccgagtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 300  
DB 910 Ataccgagtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 969  
QY 301 gggcattcctccgctgacgtcgtgacgtgacgtgacgtgacgtgacgtgacgtgacgt 360  
DB 970 Gggcattcctccgctgacgtcgtgacgtgacgtgacgtgacgtgacgtgacgtgacgt 1029  
QY 361 acaggaacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac 396  
DB 1030 Acaggaacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac 1065
```

RESULT 2
AX005790 1143 bp DNA linear PAT 24-AUG-2000
LOCUS
DEFINITION Sequence 909 from Patent WO9909186.
ACCESSION AX005790
VERSION AX005790.1 GI:9928797
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis.
Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE
1 (bases 1 to 1143)
Portnoi, D. and Guigueno, A.
Polypeptide nucleic sequences exported from mycobacteria, vectors
comprising same and uses for diagnosing and preventing tuberculosis
Patent: WO 9909186-A 909 25-FEB-1999;
JOURNAL
PORTNOI DENIS (FR); GUIGUENO AGNES (FR)
location/Qualifiers

FEATURES
source
1..1143
/organism="Mycobacterium tuberculosis"
/db_xref="taxon:1773"
4..1143
/note="SEQ ID NO 50F"

```
/note="50D"  
/codon_start=1  
/transl_table=11  
/protein_id="CAC05179.1"  
/db_xref="GI:9928798"  
/translation="KNPAARTKACTISDSTPDGCGYNSNSRRSLRMSWLLSVLAA  
VGLGLATAPQAAPALSGDREADPALDPSAMVAOVPQVAVNTKLGNNNAAGAG  
GTGIVYIDNGVLTNNHVLGATIDINAFVSGQYVGVVADVAIQLRGAGLPSAIGGV  
AVGEPVAMNGS GGTPRAVPGRVVAVGOTVQASDLSLTGAEETLNLGIOPDAIOPDSDGPPVNGIG  
VGMNTAASDNFOLSGGGGFAPIGQAMAIAGQIRSSGGSPVHIGPTAFGLGIVD  
NNGNARVORVVASPAASLIGTSDGVITAVGAPRINSATMAADLNHHPDVISVT  
WQTKSGRTGTAVTAEPRPA"
```

BASE COUNT 189 a 373 c 395 g 186 t
ORIGIN

Query Match 100.0%; Score 396; DB 6; Length 1143;
Best Local Similarity 100.0%; Pred. No. 7.4e-56;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 acggccgctgcgataacttcagctgctccaggtgaggagagatcgccattccgac 60  
DB 745 ACggccgctgcgataacttcagctgctccaggtgaggagagatcgccattccgac 804  
QY 61 gggcagcgatgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 120  
DB 805 Gggcagcgatgagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 864  
QY 121 atcgagcctacgccttcctcgcgtggtgtgtgtgcacaaacagcgagcgacga 180  
DB 865 Atcgagcctacgccttcctcgcgtggtgtgtgtgcacaaacagcgagcgacga 924  
QY 181 gtccaaacgctgctgagcgagcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240  
DB 925 Gtccaaacgctgctgagcgagcgtccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 984  
QY 241 ataccgagtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 300  
DB 985 Ataccgagtgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgagcgag 1044  
QY 301 gggcattcctccgctgacgtcgtgacgtgacgtgacgtgacgtgacgtgacgtgacgt 360  
DB 1045 Gggcattcctccgctgacgtcgtgacgtgacgtgacgtgacgtgacgtgacgtgacgt 1104  
QY 361 acaggaacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac 396  
DB 1105 Acaggaacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgacgtgac 1140
```

RESULT 3
MTC1418B 11700 bp DNA linear BCT 03-AUG-2001
LOCUS
DEFINITION Mycobacterium tuberculosis H37Rv complete genome; segment 7/162.
ACCESSION Z96071 AL123456
VERSION Z96071.1 GI:3242254
KEYWORDS
SOURCE
ORGANISM
Mycobacterium tuberculosis H37Rv.
Mycobacterium tuberculosis
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.

REFERENCE
1 (bases 1 to 11700)
Cole, S.T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C.,
Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry III, C.E.,
Tejeda, F., Badcock, K., Basham, D., Brown, D., Brown, J., Chillingworth, T.,
Connor, R., Davies, R., Devlin, K., Feltham, D., Gentles, S.,
Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., McLean, J.,
Moule, S., Murphy, L., Oliver, S., Osborne, J., Quail, M.A.,
Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, S.,
Squares, S., Squires, J., Sulston, J.E., Taylor, K., Whitehead, S. and
Barrell, B.G.
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
Nature. 393 (6685), 537-544 (1998)

MEDLINE
REFERENCE
AUTHORS
TITLE
JOURNAL98295987
2 (bases 1 to 11700)
Parkhill,J.Direct Submission
Submitted (11-JUN-1998)

Submitted on behalf of the Mycobacterium tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 20, 1998 this sequence version replaced gi:2181960.

COMMENT

Notes:

Details of M. tuberculosis sequencing at the Sanger Centre are available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/M.tuberculosis/>) CDS have been renumbered from the original cosmid submissions but the old gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes implemented in Tlpase (Krogh) supplemented with visual inspection of positional base preference in codons, especially where there is an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

FEATURES
Location/Qualifiers

source

1. 11700
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"

misc_feature

1. 1409
/note="fragment designated v031. Does not represent a physical clone"83. 88
/note="possible RBS, AAGGAG, for Rv0119"

RBS

97. 1674
/gene="fadD7"

gene

97. 1674
/gene="fadD7"
/note="Rv0119, (MTV031.13-MTC1418B.01), len: 525.fadD7, possible coenzyme A-ligase similar to 4-comurate:CoA ligase of many organisms e.g. PTU39405.1 (537 aa). Contains PS00455 putative AMP-binding domain signature. FASTA scores: gpl039405|PTU39405.1 Plus taeda xylem 4-comurate:CoA (537 aa) opt:4832-score: 526.1 E(): 8.3e-22: 28.2% identity in 440 aa overlap score is 0.896"

CDS

/codon_start=1
/transl_table=11
/product="fadD7"
/protein_id="CAB09455.1"
/db_xref="GI:3242255"
/db_xref="SPTREMBL:O07169"
/translation="MASDFGPRIDLVFAATRLPEAPALVVTADIAISHRDARLV DELAGOLTRSGILPDGDRALRWGSAEFAEVALILASRDALVPLDPAITPEORVS QAAGRVYLIDPDGPHDRAEPTTRMPLTYNAGSGSGSLSHLDAATPEPRATIS TPEGRLPDAMTETGGTGLPKMVPWTHANASSVRAITGYRLSPDAVAAPPLY HGHGILASLALASGAVSLPARQFSAHFTWDIKAVGATWYAVPTIHOILLERS ATEPSGRPAALRTIRSCSAPLTQAALALQEPFAAPVCAFGMTAEATHOVTTQIEG IDQETPVYSTGLVSTGAQIRIVSGDLPLPAGAVEIMLRGTTVAGVADPTIT ANFTDGMRLTGDLGSLAGDLSTIRGIRKELINGGEKISPEREVLGASHPMMEA AVEGVPHOLGSAVAIVPRSAAPPTREELVQFORELAEFTASQFASGLPHTA KSLDRRAVAERFGHSV"

misc_feature

610. 645
/gene="fadD7"
/note="PS00455 Putative AMP-binding domain signature, [LIVMFY].[2][STG][STAG][GIST][SPEI][SC].[PASTLVM][KR], info count = 22.0"

source

1309. 11293
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="I418B"gene
CDScomplement(1675..3819)
/gene="fusA2"
complement(1675..3819)
/gene="fusA2"
/note="Rv0120c, (MTC1418B.02c), fusA2, len: 714. Elongation factor 9, similar to e.g. EFG-ECOLI P02996 elongation factor 9 (ef-g). Escherichia coli (703 aa), fasta scores: opt:1049 z-score: 1105.5 E(): 0.32.5% identity in 717 aa overlap. Contains PS00017 ATP/GTP-binding site motif A (P-loop). Also similar to M. tuberculosis fusA, MTCY210.01 (39.1% identity in 299 aa overlap). Similar to sp|P30767|EFG_MYCLE ELONGATION FACTOR G (EF-G). (701 aa): 31.7% identity in 710 aa overlap."

misc_feature

/codon_start=1
/transl_table=11
/product="fusA2"
/protein_id="CAB09448.1"
/db_xref="GI:2181962"

gene

/translation="MADRVNASGAAAPRANGPGVRRNVLPVSGGKRTLLIEALL VAAKVLSPGSEVTEGTVCDEDEAETROQSGVLAVALADGIKVNLVDTGADFEV GELRAGLRADCALEFVIANEGYDEPTSLMOGSCOVMPRAVITKLDHARANYREA LTAQADRGKVLVLPYPSGDLIGLSOLALYERADGRRTTPRPAESDTEREAGA LIEGIIRESDESIMERYIGEGTIDTSVLIDLEKAVAGSFEPPVTPVCSSTGVTLE LIEVATRGFPSPMEHPLPEVFTTPOGPHALACDNDARPLAEVYTKTSPPYGRVSLY RVESGTRIPDTYHVSCHSFEFGGTSNTHPHDEBERGLVLFPLGKQKRPAAVY AGDICAIGKLSRAETGDTLSDAEPLVLRKWTPEPLTALIAAKATDEKLSVLG RLAADETPLRIEONQETHOVLMCMGPAHAGVLDLRLANKYGSVPTIELRPLRETF AGNAKGGRHRIKOSGGHGYCVDIEVEPIEGSCGEFLDKVVGAVPAPQFIPLNKEF VRAOMDKGVHAGVPVVDIRVTLIDGKRHSYDSDFAFOMAGALALAEAAATVILILE PIDELISLVDPDFFGAVLIDSLSSRGVIGTETAGHDRTYIKAEVQVELTFRALDLR SLHGAASFTRSFARYPEMPESAARAKNAG"

CDS

complement(3709..3732)
/gene="fusA2"
/note="PS00017 ATP/GTP-binding site motif A (P-loop)"
complement(3956..4390)
/gene="Rv0121c"
complement(3956..4390)
/gene="Rv0121c"
/note="Rv0121c, (MTC1418B.03c), unknown, len: 144 aa"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0121c"
/protein_id="CAB09449.1"
/db_xref="GI:2181963"
/db_xref="SPTREMBL:O07171"
/translation="MGSEFDPKLRPAOSPARLATSPDGTGPHLVPAVFAAGARPAA TGAIVTYAVDAKARKTQRLRLANLEHNRASVLVDSYADPTQLMWRADGVAAIH RDEGVRAATRLRAKTAQYOSVPLNPVLAIVORRASHNA"
complement(4399..4401)
/note="possible RBS upstream of Rv0121c"
4332..4535
/note="possible RBS upstream of Rv0122"
4539..4907
/gene="Rv0122"
/note="Rv0122, (MTC1418B.04), unknown, len: 133 aa"

RBS

RBS

gene
CDS/codon_start=1
/transl_table=11
/product="hypothetical protein Rv0122"
/protein_id="CAB09450.1"
/db_xref="GI:2181964"
/db_xref="SPTREMBL:O07172"
/translation="MAGSVSAAGIGVAGLNTETNRDQCYRVERPTVDALHPREYR HTRGVQVRYTRAKRHRYVAKHRIVAAHRCGVVIOEDSLVYGGSDISGRLTVEVA VEADGDLITTHAMPKEMKR"
4890..4895
/gene="Rv0122"
/note="possible RBS upstream of Rv0123"
4904..5272
/gene="Rv0123"
4904..5272

RBS

gene

CDS

```

/ gene="RV0123"
/ note="RV0123, (MTC1418B.05), unknown, len: 133 aa"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein RV0123"
/ protein_id="CA09451.1"
/ db_xref="GI:2181965"
/ db_xref="SPTREMBL:007173"
/ translation="MTKKPRNPADYVIGDDVEVSDVLDKOEYVVDGERLDERVEOM
ASESLRLAREANLIPGKSLSGSAHSAVQYVVSATKAKKELARSKMSVSKL
LRPVDPERETGRILPRR"
RBS
5570..5574
/ note="possible RBS upstream of RV0124"
5581..7044
/ gene="PE_PGRS"
5581..7044
/ gene="PE_PGRS"
CDS
5581..7044
/ gene="PE_PGRS"
/ note="RV0124, (MTC1418B.06), len: 487. Member of the
PGRS family of glycine-rich proteins, highly similar to
eg. Y0DP_MYCTU_Q50615 hypothetical glycine-rich 40.8 kd
protein (498 aa), fasta scores: opt: 1730 z-score: 1201.5
E(): 0, 60.7% identity in 504 aa overlap"
/ codon_start=1
/ transl_table=11
/ product="PE_PGRS"
/ protein_id="CA09452.1"
/ db_xref="GI:2181966"
/ db_xref="SPTREMBL:007174"
/ translation="MSFVSVAPEYVVAATDIAGISASIAANAAAAAPTAVIAGA
DEVSAITALEFSGHQAIVQALSNQAAAFHOQFVOTLGGAGATRAAEQVEOLLAI
NAPTOALGRPLINGADGAGTGQAGAGILYNGNGSGSAGAGSAGAGPGLI
Query Match 100.0%; Score 396; DB 1; Length 11700;
Best Local Similarity 100.0%; Pred. No. 5.1e-56;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

SOURCE
ORGANISM
Mycobacterium tuberculosis CDC1551.
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium tuberculosis complex.
1 (bases 1 to 14029)
REFERENCE
AUTHORS
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE
Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains
JOURNAL
Unpublished
2 (bases 1 to 14029)
REFERENCE
AUTHORS
Fleischmann, R.D., Alland, D., Eisen, J.A., Carpenter, L., White, O.,
Peterson, J., Deboy, R., Dodson, R., Gwinn, M., Haft, D., Hickey, E.,
Kolony, J.F., Nelson, W.C., Umayam, L.A., Ermolaeva, M.,
Salberg, S.L., Delcher, A., Utterback, T., Weidman, J., Khouri, H.,
Gill, J., Mikula, A. and Bishai, W.
TITLE
Direct Submission
JOURNAL
Submitted (25-Apr-2001) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
FEATURES
Source
1..14029
/ organism="Mycobacterium tuberculosis CDC1551"
/ strain="CDC1551"
/ db_xref="taxon:83331"
/ note="clinical strain"
complement(99..533)
/ gene="MT0129"
complement(99..533)
/ gene="MT0129"
/ note="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AKK4353.1"
/ db_xref="GI:13879611"
/ translation="MGEFDPKLRFAQSFVARLSTPDGTPLVPPVAFGARRPAA
TGADIVYAVDAKRTQRLRLANLHNPPASVLYSDADMTOLMWWVRADGVAIIR
RDEVMRAAYRLRLAKYTOYOSVPLNGVIAIAVQRMASRHA"
complement(643..1062)
/ gene="MT0130"
complement(643..1062)
/ gene="MT0130"
/ note="identified by Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="hypothetical protein"
/ protein_id="AKK4354.1"
/ db_xref="GI:13879612"
/ translation="MASSSSPLRPHCVSDQVYVGPDPDGLKTAARTAIYVORA
IFLNDNTRVAHSGDAVAGHAVLFGVRRDDPLHASSMYSITLKCOSVNGRPDAIA
LVSVRCHVDYPTDSCGGRDGPGLPCAPLDYHRHH"
1047..1415
/ gene="MT0131"
1047..1415
/ gene="MT0131"
/ note="identified by match to protein family HMM"
/ codon_start=1
/ transl_table=11
/ product="DNA-binding protein, CopG family"
/ protein_id="AKK4355.1"
/ db_xref="GI:13879613"
/ translation="MTKKPRNPADYVIGDDVEVSDVLDKOEYVVDGERLDERVEOM
ASESLRLAREANLIPGKSLSGSAHSAVQYVVSATKAKKELARSKMSVSKL
LRPVDPERETGRILPRR"
1682..3367
/ gene="MT0132"
1682..3367
/ gene="MT0132"
/ note="similar to GB:AL123456; identified by sequence

```


QY 301 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 360
 |||||||
 Db 4488 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 4547
 |||||||
 QY 361 acaggaacgtgacatctgacgaggaaccccgcc 396
 |||||||
 Db 4548 acaggaacgtgacatctgacgaggaaccccgcc 4583
 |||||||

RESULT 5

AR169152 447 bp DNA linear PAT 17-DEC-2001
 LOCUS AR169152
 DEFINITION Sequence 4 from patent US 6290969.
 ACCESSION AR169152
 VERSION AR169152.1 GI:17906927
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 447)
 AUTHORS Reed,S.G., Skelky,Y.A.W., Dillon,D.C., Campos-Neto,A., Houghton,R.,
 Vedvik,T.S. and Twardzik,D.R.
 TITLE Compounds and methods for immunotherapy and diagnosis of
 tuberculosis
 JOURNAL Patent: US 6290969-A 4 18-SEP-2001;
 FEATURES Location/Qualifiers
 source 1..447
 BASE COUNT 79 a 146 c 149 g 72 t 1 others
 ORIGIN

Query Match 99.6%; Score 394.4; DB 6; Length 447;
 Best Local Similarity 99.7%; Pred. No. 1.6e-55;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggcgcgctcgatatactccagctgctccagggcgagggatcgccattccgac 60
 |||||||
 Db 11 acggcgcgctcgatatactccagctgctccagggcgagggatcgccattccgac 70
 |||||||
 QY 61 gggcagcgatgacgacgacgacgacgacgacgacgacgacgacgacgacgac 120
 |||||||
 Db 71 gggcagcgatgacgacgacgacgacgacgacgacgacgacgacgacgacgac 130
 |||||||
 QY 121 atcgagcgctacgacgacgacgacgacgacgacgacgacgacgacgacgac 180
 |||||||
 Db 131 atcgagcgctacgacgacgacgacgacgacgacgacgacgacgacgacgac 190
 |||||||
 QY 181 gtccaaacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 240
 |||||||
 Db 191 gtccaaacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 250
 |||||||
 QY 241 atcacgcgctgacgacgacgacgacgacgacgacgacgacgacgacgacgac 300
 |||||||
 Db 251 atcacgcgctgacgacgacgacgacgacgacgacgacgacgacgacgacgac 310
 |||||||
 QY 301 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 360
 |||||||
 Db 311 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 370
 |||||||
 QY 361 acaggaacgtgacatctgacgaggaaccccgcc 396
 |||||||
 Db 371 acaggaacgtgacatctgacgaggaaccccgcc 406
 |||||||

RESULT 6
 BD006325 447 bp DNA linear PAT 31-JAN-2002
 LOCUS BD006325
 DEFINITION Compounds and methods for diagnosis of Tuberculosis.
 ACCESSION BD006325
 VERSION BD006325.1 GI:18634696
 KEYWORDS JP 2001500383-A/4.
 SOURCE unidentified.

ORGANISM unidentified
 unclassified.
 REFERENCE 1 (bases 1 to 447)
 AUTHORS Reed,S.G., Skelky,Y.A.W., Dillon,D.C., Neto,A.C., Houghton,R.,
 Vedvik,T.S., Twardzik,D.R. and Lodes,M.J.
 TITLE Compounds and methods for diagnosis of Tuberculosis
 JOURNAL Patent: JP 2001500383-A 4 16-JAN-2001;
 CORIXA CORP
 OS Unidentified
 PN JP 2001500383-A/4
 PD 16-JAN-2001
 PE 07-OCT-1997 JP 1998518432
 PR 11-OCT-1996 US 08/729622,13-MAR-1997 US 08/818111 PI
 STEVEN G REED, YASIR A W SKEIKY, DAVIN C DILLON, PI ANTONIO CAMPOS
 NETO,
 PI RAYMOND HOUGHTON, THOMAS S VEDVIK, DANIEL R TWARDZIK, PI
 MICHAEL J LODES
 PC C12N15/31,C07K14/35,C07K16/12,C12Q1/68,C12N15/62,G01N33/53 CC
 Strandedness: Single;
 CC Topology: Linear;
 FH Key Location/Qualifiers
 FT source 1..447
 /organism='unidentified'.
 /db_xref='taxon:32644'
 BASE COUNT 79 a 146 c 149 g 72 t 1 others
 ORIGIN

Query Match 99.6%; Score 394.4; DB 6; Length 447;
 Best Local Similarity 99.7%; Pred. No. 1.6e-55;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggcgcgctcgatatactccagctgctccagggcgagggatcgccattccgac 60
 |||||||
 Db 11 acggcgcgctcgatatactccagctgctccagggcgagggatcgccattccgac 70
 |||||||
 QY 61 gggcagcgatgacgacgacgacgacgacgacgacgacgacgacgacgacgac 120
 |||||||
 Db 71 gggcagcgatgacgacgacgacgacgacgacgacgacgacgacgacgacgac 130
 |||||||
 QY 121 atcgagcgctacgacgacgacgacgacgacgacgacgacgacgacgacgac 180
 |||||||
 Db 131 atcgagcgctacgacgacgacgacgacgacgacgacgacgacgacgacgac 190
 |||||||
 QY 181 gtccaaacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 240
 |||||||
 Db 191 gtccaaacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 250
 |||||||
 QY 241 atcacgcgctgacgacgacgacgacgacgacgacgacgacgacgacgacgac 300
 |||||||
 Db 251 atcacgcgctgacgacgacgacgacgacgacgacgacgacgacgacgacgac 310
 |||||||
 QY 301 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 360
 |||||||
 Db 311 gggcatcaccggtgacatctcgtgacatgcaaaccaagtcggcgacacgct 370
 |||||||
 QY 361 acaggaacgtgacatctgacgaggaaccccgcc 396
 |||||||
 Db 371 acaggaacgtgacatctgacgaggaaccccgcc 406
 |||||||

RESULT 7
 BD006445 447 bp DNA linear PAT 31-JAN-2002
 LOCUS BD006445
 DEFINITION Compounds and methods for immunotherapy and diagnosis of
 Tuberculosis.
 ACCESSION BD006445
 VERSION BD006445.1 GI:18634816
 KEYWORDS JP 2001501832-A/4.
 SOURCE unidentified.

ORGANISM	unidentified unclassified
REFERENCE	1 (phases 1 to 447)
AUTHORS	Reed,S.G., Skelzy,I.A.W., Dillon,D.C., Neto,A.C., Houghton,R., Vedvick,T.S., Twardzik,D.R. and Lodes,M.J.
TITLE	Compounds and methods for immunotherapy and diagnosis of
JOURNAL	Patent: JP 2001501832-A 4 13-FEB-2001; CORIX CORP
COMMENT	OS Unidentified PN JP 2001501832-A/4 PD 13-FEB-2001 PP 07-OCT-1997 JP 1998518456 PR 11-OCT-1996 US 08/730510,13-MAR-1997 US 08/818112 PI SEVEN G REED,YASIR A W SKEITZ,DAVIN C DILLON, PI ANTONIO CAMPOS NETO, PI RAYMOND HOUGHTON,THOMAS S VEDVICK,DANIEL R TWARDZIK, PI MICHAEL J LODES PC C12N15/31,C07K14/35,A61K39/04,A61K48/00,A61K49/00,C12N15/62, PC C07K19/00 PC G01N33/50,G01N33/60,G01N33/569,C12N1/19,C12N1/20,C12N1/21, PC C12N5/10// PC (C12N1/21,C12R1:19) CC Strandedness: Single; CC Topology: Linear; FH Key Location/Qualifiers FT source 1. 447 FT /organism='Unidentified', Location/Qualifiers 1. 447 /organism='unidentified' /db_xref='taxon:32644'
BASE COUNT	79 a 146 c 149 g 72 t 1 others
ORIGIN	
FEATURES	
Source	
Query Match	99.68; Score 394.4; DB: 6; Length 447;
Best Local Similarity	99.78; Pred. No. 1.6e-55;
Matches 395; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	1 acggcgcgcgtccgaatactccaagctgtgcccaaggctggcaggaattccgcattccgaatc 60
DB	11 ACGGCGCGCTCCGATTAACCTTCACGCTGTCCACGAGGTGGGAGGATTCGCCATTCGATC 70
QY	61 gggcagcgatgagcgcatctgcggcgccagatccgattgggtgggggtaccaccacgttat 120
DB	71 GGGCAGGCGATGGCATGTGGGCGCCAGATCCGATGGGGTGGGCTCACCCACCTTCAT 130
QY	121 atcggcgctaccgcgctctcctcgcgcttgggtgtgtgtgcgaacaacaaaggaacgycgcacga 180
DB	131 ATCGGCGCTTACCGCGCTTCTCTCGGCTTGGGTGTGTGTGCACAAACGCGACCGCCACACA 190
QY	181 gtcccaacgcgtgttcgggaagcgctccggcgccgaattccgcattccaccgcggagcgtg 240
DB	191 GTCCACACCGGTGTGTGGGAGCGCTCCGGCGGCAATCTTCGGCATCTCCACCGCCACGTG 250
QY	241 atcacgcggttcgacgagcgcttcgaatcaactcggcaccacgcgaatggcggacgcgcttaac 300
DB	251 ATCACCGCGGTTCGACGCGCTCCGATTAACCTTCGCGACCGCATGCGCGACCGCTTAAAC 310
QY	301 gggcattatcccggtggaagtcacttcggttgcattggaaacaaagtcggcgacgcgct 360
DB	311 GGGCATTCATCCCGGTGAGCTCATCTTCGGTGAACCTGGCAACCAAGTCCGGCGGACGCGCT 370
QY	361 acaggaacgttgacattggccgaggaagcccccgcgc 396
DB	371 ACAGGGAACGTGTACATGTGGCCGAGGAGACCCCGGCC 406
RESULT	8
LOCUS	AR169165 1872 bp DNA linear PAT 17-DEC-2001
DEFINITION	Sequence 17 from patent US 6230969.
ACCESSION	AR169165

VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE AUTHORS	TITLE	JOURNAL FEATURES	BASE COUNT	ORIGIN
ARI69165.1	GI:17906947	Unknown.	Unknown.	Unclassified. 1 (bases 1 to 1872)	Reed,S.G., Skeiky,Y.A.W., Dillon,D.C., Campos-Neto,A., Houghton,R., Vedvick,T.S. and Twardzik,D.R. Compounds and methods for immunotherapy and diagnosis of tuberculosis Patent: US 6290969-A 17 18-SEP-2001;	Location/Qualifiers .1.1872	318 a 616 c 604 g 332 t 2 others	
Query Match	Best Local Similarity	Matches 393;	Conservative	98.8%; Score 391.2; DB 6: 99.2%; Pred. No. 4,2e-55;	Length 1872; Mismatches 3;	Indels 0;	Gaps 0;	
OY	I	acggcgcgctcgataacttcagctgtcccaagggtggcaggatcgcattccgcatc 60						
Db	758	ACGGCCCGCTCCATTAACATTCCAGCTGTCCCAGGCTGGCGAGGATTCGCCATTCCGCATC 817						
OY	61	ggcgaaagcagtgcagtcgcggcggaagtcacgatccgatgggtgggggtaccacacgttcat 120						
Db	818	GGCGAAGCGATGGCGCATCCGGGGCCAATCGATGGGTGGGGGGTCAACCACCGTTCAAT 877						
OY	121	atcggcgctaacgccttcctcgcgcttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180						
Db	878	ATCCGGCGCTACCCCTTCCTCGGCTGGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 937						
OY	181	gtccaaacgcgtgtgtgtggagagcgtccggcgcggaagtctcgcgcatctcacacgcgcagctg 240						
Db	938	GTCACAACGCTGTGTGTCGAAGAAGCGTCCGGGGCGAAATCTTCGGCATCTCACCGGGAGAGT 997						
OY	241	atcacgcggttcgaagcgcgctccgatalcaacttcgcccacgcgcatgctggcgagcgcttaac 300						
Db	998	ATCACCCCGTGTCACAGCGCGCTCGAATAACTCGGCACCGCGATGGCGAGCGCGTTAAC 1057						
OY	301	gggcatactcccggttaagctcatctcgtgtgcgttcgtagccttgccaacaagaatcgggcgagcgtg 360						
Db	1058	GGGCATCATCCCGGTGACATCTCATCTCGTGTAACATGGCAACCAAGTGGGGGGGACGCGT 1117						
OY	361	acaaggaacgtgacatctgcccgaaggagcccccgcc 396						
Db	1118	ACAGGGAACGTGACATTGGCCGAGGAGCCCCCGCGCC 1153						
RESULT 9								
BD006338			1872 bp	DNA linear	PAT 31-JAN-2001			
LOCUS	BD006338							
DEFINITION	Compounds and methods for diagnosis of Tuberculosis.							
ACCESSION	BD006338							
VERSION	BD006338.1							
KEYWORDS	JP 2001500383-A/17.							
SOURCE	unidentified.							
ORGANISM	unclassified.							
REFERENCE	1 (bases 1 to 1872)							
AUTHORS	Reed,S.G., Skeiky,Y.A.W., Dillon,D.C., Neto,A.C., Houghton,R., Vedvick,T.S., Twardzik,D.R. and Lodes,M.J.							
TITLE	Compounds and methods for diagnosis of Tuberculosis							
JOURNAL	Patent: JP 2001500383-A 17 16-JAN-2001;							
COMMENT	CORIXA CORP							
	OS Unidentified							
	PN JP 2001500383-A/17							
	PD 16-JAN-2001							
	PF 07-OCT-1997 US 08/722962,13-MAR-1997 US 08/818111 PI							
	PR 11-OCT-1996 US 08/722962,13-MAR-1997 US 08/818111 PI							
	STEVEN G REED, YASIR A W SKEIKY, DAVIN C DILLON, PI ANTONIO CAMPOS							

RESULT	11				
LOCUS	AX201049				
DEFINITION	Sequence 679 from Patent WO0151633.	675 bp	DNA	linear	PAT 29-AUG-2001
ACCESSION	AX201049				
VERSION	AX201049.1				
KEYWORDS	GI:15390857				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Ekahyوتا; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 675)				
AUTHORS	Xu, J., Dillon, D.C., Micham, J. L., Harlocker, S. L., Jiang, Y., Reed, S. C., Kalos, M. D., Fanger, G. N., Day, C. H., Ketter, R. W., Stolz, J. A., Skelky, J. A., Wang, A. and Weagater, M. W.				
TITLE	Compositions and methods for the therapy and diagnosis of prostate cancer				
JOURNAL	Patent: WO 0151633-A 679				
	CORIXA CORPORATION (US)	19-JUL-2001;			


```

QY 241 atcacccggtcagcgcgtcccatcaactcgcacccgcatgctgcagcgcgttaac 300
    |||||||
DB 250 ATACCGCGGTGACGCGCGTCCGATCAATCGGCGCATGGCGGACGCGCTTAAC 309
    |||||||
QY 301 gggcatcctccggtgacgtcattcgtggtgacctgcaacaaagtcggcgacgcgt 360
    |||||||
DB 310 GGGCATCATCCCGGTGACGTATCTCGTGACCTGGCAAAAGTCGGCGGACGCGT 369
    |||||||
QY 361 acagggagctgacatgctgctgcagagggagcccgcc 396
    |||||||
DB 370 ACAGGGAACCTGACATTGGCCGAGGAGACCCCGGCC 405
    |||||||

RESULT 14
AX201061 915 bp DNA linear PAT 29-AUG-2001
LOCUS Sequence 691 from Patent WO0151633.
DEFINITION AX201061
ACCESSION AX201061
VERSION AX201061.1 GI:15390868
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 915)
Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Reed,S.G., Kalos,M.D., Fanger,G.R., Day,C.H., Ketter,M.W.,
Stolk,J.A., Skeiky,Y.A., Wang,A., and Meagher,W.J.
Compositions and methods for the therapy and diagnosis of prostate
cancer
Patent: WO 0151633-A 691 19-JUN-2001;
CORIXA CORPORATION (US)
FEATURES
source
1..915
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 167 a 280 c 284 g 184 t
ORIGIN

Query Match 89.4%; Score 354; DB 6; Length 915;
Best Local Similarity 95.7%; Pred. No. 5.7e-49;
Matches 379; Conservative 0; Mismatches 5; Indels 12; Gaps 1;

QY 1 acggcgcgtcgcataactccagctgtcccaagggtggcagagatccgcatccgac 60
    |||||||
DB 22 ACGGCGCGTCCGATTAATCTCCAGCTGTCCAGGCGGCGCATTCGCGATC 81
    |||||||
QY 61 gggcagcgatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 120
    |||||||
DB 82 GGGCAGGCGATGGCGATCGCGGCGGCGCATC-----AAGCTTCCACCGCTTCA 129
    |||||||
QY 121 atcggtcctacccgtcctcctcctcctcctggtgtgtgtgtgtgtgtgtgtgtgt 180
    |||||||
DB 130 ATCGGGCGCTACCGCTTCTCTCGGCTTGGTGTGTGTGTGTGTGTGTGTGTGTGT 189
    |||||||
QY 181 gtccaaagctgtgtcggagagctccggcgcaagtcctgcatctccacggagagtg 240
    |||||||
DB 190 GTCCAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 249
    |||||||
QY 241 atcacccggtcagcgcgtcccatcaactcgcacccgcatgctgcagcgcgttaac 300
    |||||||
DB 250 ATACCGCGGTGACGCGCGTCCGATCAATCGGCGCATGGCGGACGCGCTTAAC 309
    |||||||
QY 301 gggcatcctccggtgacgtcattcgtggtgacctgcaacaaagtcggcgacgcgt 360
    |||||||
DB 310 GGGCATCATCCCGGTGACGTATCTCGTGACCTGGCAAAAGTCGGCGGACGCGT 369
    |||||||
QY 361 acagggagctgacatgctgctgcagagggagcccgcc 396
    |||||||
DB 370 ACAGGGAACCTGACATTGGCCGAGGAGACCCCGGCC 405
    |||||||

```

```

RESULT 15
AX267860 915 bp DNA linear PAT 26-OCT-2001
LOCUS Sequence 834 from Patent WO0173032.
DEFINITION AX267860
ACCESSION AX267860
VERSION AX267860.1 GI:16516503
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (sites)
Xu,J., Dillon,D.C., Mitcham,J.L., Harlocker,S.L., Jiang,Y.,
Kalos,M.D., Fanger,G.R., Ketter,M.W., Stolk,J.A., Day,C.H.,
Vedvick,T.S., Carter,D., Li,S.X., Wang,A., Skeiky,Y.A., Hepler,W.T.
and Henderson,R.A.
Compositions and methods for the therapy and diagnosis of prostate
cancer
Patent: WO 0173032-A 834 04-OCT-2001;
CORIXA CORPORATION (US)
FEATURES
source
1..915
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 167 a 280 c 284 g 184 t
ORIGIN

Query Match 89.4%; Score 354; DB 6; Length 915;
Best Local Similarity 95.7%; Pred. No. 5.7e-49;
Matches 379; Conservative 0; Mismatches 5; Indels 12; Gaps 1;

QY 1 acggcgcgtcgcataactccagctgtcccaagggtggcagagatccgcatccgac 60
    |||||||
DB 22 ACGGCGCGTCCGATTAATCTCCAGCTGTCCAGGCGGCGCATTCGCGATC 81
    |||||||
QY 61 gggcagcgatgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 120
    |||||||
DB 82 GGGCAGGCGATGGCGATCGCGGCGGCGCATC-----AAGCTTCCACCGCTTCA 129
    |||||||
QY 121 atcggtcctacccgtcctcctcctcctcctggtgtgtgtgtgtgtgtgtgtgtgt 180
    |||||||
DB 130 ATCGGGCGCTACCGCTTCTCTCGGCTTGGTGTGTGTGTGTGTGTGTGTGTGTGT 189
    |||||||
QY 181 gtccaaagctgtgtcggagagctccggcgcaagtcctgcatctccacggagagtg 240
    |||||||
DB 190 GTCCAAAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 249
    |||||||
QY 241 atcacccggtcagcgcgtcccatcaactcgcacccgcatgctgcagcgcgttaac 300
    |||||||
DB 250 ATACCGCGGTGACGCGCGTCCGATCAATCGGCGCATGGCGGACGCGCTTAAC 309
    |||||||
QY 301 gggcatcctccggtgacgtcattcgtggtgacctgcaacaaagtcggcgacgcgt 360
    |||||||
DB 310 GGGCATCATCCCGGTGACGTATCTCGTGACCTGGCAAAAGTCGGCGGACGCGT 369
    |||||||
QY 361 acagggagctgacatgctgctgcagagggagcccgcc 396
    |||||||
DB 370 ACAGGGAACCTGACATTGGCCGAGGAGACCCCGGCC 405
    |||||||

```

Search completed: August 6, 2002, 11:05:22
 Job time: 7707 sec

PT tuberculosis -

XX Example: Fig 13A-B; 83pp; English.

XX This DNA sequence includes a coding region for a recombinant
PS Mycobacterium tuberculosis bi-antigen fusion protein (see AAY32071),
CC termed Mtb24, composed of the antigens Ral2 and DPP. The
CC DNA is useful for the recombinant production of the fusion protein.
CC Coding sequences for the antigens were modified by PCR in order
CC to facilitate their fusion and subsequent expression of the fusion
CC protein, and then ligated. The invention provides fusion proteins
CC (see AAY32059-71) containing at least 2 M. tuberculosis antigens. The
CC new fusion proteins and polynucleotides encoding them are useful as
CC vaccines for preventing tuberculosis (claimed), for diagnosis (via
CC in vitro assays or intradermal skin tests for detection of anti-M.
CC tuberculosis antibodies), monitoring of disease progression, and
CC treatment of tuberculosis. They are more effective immunogens than
CC mixtures of the individual protein components.

XX Sequence 702 BP; 127 A; 229 C; 225 G; 121 T; 0 other;

Query Match

100.0%; Score 396; DB 20; Length 702;

Best Local Similarity 100.0%; Pred. No. 2e-85;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acggcgcgttcgatacttccagctgttccaggttgagcaggttcgcatccgac 60

DB 25 acggcgcgttcgatacttccagctgttccaggttgagcaggttcgcatccgac 84

QY 61 gggcagcgtatgagatcgagcagcagatcgatcgatggtgggggttcaaccacgttcat 120

DB 85 gggcagcgtatgagatcgagcagcagatcgatcgatggtgggggttcaaccacgttcat 144

QY 121 atcgagcgtacccgcttcctcgttgggtgtgttcgcaaaaacagcagcagcagcaga 180

DB 145 atcgagcgtacccgcttcctcgttgggtgtgttcgcaaaaacagcagcagcagcaga 204

QY 181 gtccaaacgctgttcgagcagcgtccgagcagatctcgatctcaaccgagcagctg 240

DB 205 gtccaaacgctgttcgagcagcgtccgagcagatctcgatctcaaccgagcagctg 264

QY 241 ataccgcggttcgagcagcgtccgatactcgcaaccgagatggcagcagcagcttaac 300

DB 265 ataccgcggttcgagcagcgtccgatactcgcaaccgagatggcagcagcagcttaac 324

QY 301 gggcgtatcccggtgagctcatctcggtgacatggcaaaccaagtccggcgagcagct 360

DB 325 gggcgtatcccggtgagctcatctcggtgacatggcaaaccaagtccggcgagcagct 384

QY 361 acagggagcgtgacatggcagcagaggaaccccgagc 396

DB 385 acagggagcgtgacatggcagcagaggaaccccgagc 420

RESULT 2

AAK34251

ID AAK34251 standard; DNA; 1068 BP.

XX AAK34251;

XX 06-JUL-1999 (first entry)

DE Mycobacterium species nucleic acid sequence 500.

KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;

XX hybridisation; detection; vaccine; immunisation; infection; ss.

OS Mycobacterium sp.

XX WO9909186-A2.

XX 25-FEB-1999.

XX 14-AUG-1998; 98WO-FR01813.

XX 11-SEP-1997; 97FR-0011325.

XX 14-AUG-1997; 97FR-0010404.

XX (INSP) INSP PASTEUR.

XX Gicquel B, Lm EM, Pelletier V, Portnoi D, Goguet de la Salmoniere Y;

XX Guigueno A;

XX WPI; 1999-181045/15.

XX P-P-SDB; AAY05000.

PT Mycobacterial DNA vectors containing reporter constructs - for

PT identifying coding or promoter sequences involved in

PT infection-associated protein expression

XX Claim 22; Fig 50D; 309pp; French.

XX Sequences AAK34001-X34252 represent nucleic acids encoding secreted
CC proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods
CC for detecting and identifying mycobacteria, especially belonging to
CC the M. tuberculosis complex. The encoded proteins can be used in
CC vaccines for immunisation against a bacterial or viral infection.

XX Sequence 1068 BP; 169 A; 352 C; 375 G; 172 T; 0 other;

Query Match

100.0%; Score 396; DB 20; Length 1068;

Best Local Similarity 100.0%; Pred. No. 2.1e-85;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 acggcgcgttcgatacttccagctgttccaggttgagcaggttcgcatccgac 60

DB 670 acggcgcgttcgatacttccagctgttccaggttgagcaggttcgcatccgac 729

QY 61 gggcagcgtatgagatcgagcagcagatcgatcgatggtgggggttcaaccacgttcat 120

DB 730 gggcagcgtatgagatcgagcagcagatcgatcgatggtgggggttcaaccacgttcat 789

QY 121 atcgagcgtacccgcttcctcgttgggtgtgttcgcaaaaacagcagcagcagcaga 180

DB 790 atcgagcgtacccgcttcctcgttgggtgtgttcgcaaaaacagcagcagcagcaga 849

QY 181 gtccaaacgctgttcgagcagcgtccgagcagatctcgatctcaaccgagcagctg 240

DB 850 gtccaaacgctgttcgagcagcgtccgagcagatctcgatctcaaccgagcagctg 909

QY 241 ataccgcggttcgagcagcgtccgatactcgcaaccgagatggcagcagcagcttaac 300

DB 910 ataccgcggttcgagcagcgtccgatactcgcaaccgagatggcagcagcagcttaac 969

QY 301 gggcgtatcccggtgagctcatctcggtgacatggcaaaccaagtccggcgagcagct 360

DB 970 gggcgtatcccggtgagctcatctcggtgacatggcaaaccaagtccggcgagcagct 1029

QY 361 acagggagcgtgacatggcagcagaggaaccccgagc 396

DB 1030 acagggagcgtgacatggcagcagaggaaccccgagc 1065

RESULT 3

AAK34252

ID AAK34252 standard; DNA; 1143 BP.

XX AAK34252;

XX 06-JUL-1999 (first entry)

DE Mycobacterium species nucleic acid sequence 500.

Secreted protein; Mycobacterium; primer; PCR; amplification; probe; hybridisation; detection; vaccine; immunisation; infection; ss.

Mycobacterium sp.

W09909186-A2. *Compiling*

25-FEB-1999.

14-AUG-1998; 98WO-FR01813.

11-SEP-1997; 97FR-0011325.
14-AUG-1997; 97FR-0010404.

(INSP) INST PASTEUR.

Guiguel B, Lim EM, Pelletier V, Portnoi D, Goguet de la Salmoniere Y;
Guigueno A;

WPI; 1999-181045/15.
P-PSDB; AAY04830.

Mycobacterial DNA vectors containing reporter constructs - for
identifying coding or promoter sequences involved in
infection-associated protein expression

Claim 22; Fig 50F; 309pp; French.

Sequences AAX34001-X34252 represent nucleic acids encoding secreted
proteins from various Mycobacterium species microorganisms. The
CC nucleotide sequences can be used as primers and probes for methods
for detecting and identifying mycobacteria, especially belonging to
the M. tuberculosis complex. The encoded proteins can be used in
vaccines for immunisation against a bacterial or viral infection.

Sequence 1143 BP; 189 A; 373 C; 395 G; 186 T; 0 other;

Query Match Best Local Similarity 100.0%; Score 396; DB 20; Length 1143;

Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 acggcgccgctccgataactccagctccagagtgaggaggaattccgcatccgcatc 60
|||||
745 acggcgccgctccgataactccagctccagagtgaggaggaattccgcatccgcatc 804
61 gggcaggcgatgagatcgccggcgagatcgatcggtgggggttaaccacggttat 120
|||||
805 gggcaggcgatgagatcgccggcgagatcgatcggtgggggttaaccacggttat 864
121 atcgggctaccgctctccgctggtgtgtgtcgaacaacaaggcagcgacaga 180
|||||
865 atcgggctaccgctctccgctggtgtgtgtcgaacaacaaggcagcgacaga 924
181 gtccaaagcggt 240
|||||
925 gtccaaagcggt 984
241 atcaccgagctgacgagcgctccgataactccgacacgacgacgacgacgacgac 300
|||||
985 atcaccgagctgacgagcgctccgataactccgacacgacgacgacgacgacgac 1044
301 gggcagatcccggtgagatcattcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 360
|||||
1045 gggcagatcccggtgagatcattcctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1104
361 acagggaacgtgacatctggccgaggaagaccccgagcc 396
|||||
1105 acagggaacgtgacatctggccgaggaagaccccgagcc 1140

RESULT 4
AAX20194

ID AAX20194 standard; DNA; 2287 BP.

AC AAX20194;

DE 17-JAN-2000 (first entry)

DE Mycobacterium tuberculosis antigen fusion protein Mtub32A DNA.

XX Tuberculosis; antigen; fusion protein; Mtub32A; Ra12; Tbh9; Ra35;

XX diagnosis; therapy; vaccine; immunogen; ss.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT CDS 42..2231

FT /tag= a

PN W09951748-A2.

PD 14-OCT-1999.

PE 07-APR-1999; 99MO-US07717.

PR 07-APR-1998; 98US-0056556.

PR 30-DEC-1998; 98US-0223040.

PA (CORI-) CORIXA CORP.

PI Skeiky VAW, Alderson M, Campos-Neto A;

DR WPI; 1999-601610/51.

DR P-PSDB; AAY32059.

PT New fusion proteins useful for diagnosis, prevention and treatment of

PT tuberculosis -

PS Example; Fig 1A-B; 83pp; English.

XX This DNA sequence includes a coding region for a recombinant

CC Mycobacterium tuberculosis tri-antigen fusion protein (see AAY32059),

CC termed Mtub32A, composed of the antigens Ra12, Tbh9 and Ra35. The

CC DNA is useful for the recombinant production of the fusion protein.

CC Coding sequences for the antigens were modified by PCR in order

CC to facilitate their fusion and subsequent expression of the fusion

CC protein. 3 Coding sequences for Ra12, Tbh9 and Ra35 were ligated

CC to encode Mtub32A. The invention provides fusion proteins (see

CC AAY32059-71) containing at least 2 M. tuberculosis antigens. The new

CC fusion proteins and polynucleotides encoding them are useful as

CC vaccines for preventing tuberculosis (claimed), for diagnosis (via

CC in vitro assays or intradermal skin tests for detection of anti-M.

CC tuberculosis antibodies), monitoring of disease progression, and

CC treatment of tuberculosis. They are more effective immunogens than

CC mixtures of the individual protein components.

SO Sequence 2287 BP; 374 A; 723 C; 805 G; 382 T; 3 other;

Query Match Best Local Similarity 100.0%; Score 396; DB 20; Length 2287;

Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 acggcgccgctccgataactccagctccagagtgaggaggaattccgcatccgcatc 60
|||||
63 acggcgccgctccgataactccagctccagagtgaggaggaattccgcatccgcatc 122
61 gggcaggcgatgagatcgccggcgagatcgatcggtgggggttaaccacggttat 120
|||||
123 gggcaggcgatgagatcgccggcgagatcgatcggtgggggttaaccacggttat 182
121 atcgggctaccgctctccgctggt 180
|||||
183 atcgggctaccgctctccgctggt 242

QY 181 gtccacgcgtgtgctggagcgctccgcgcgaagtcttcgcatccacggcgagctg 240
Db 243 gtccacgcgtgtgctggagcgctccgcgcgaagtcttcgcatccacggcgagctg 302
QY 241 atcacgcggttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 300
Db 303 atcacgcggttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 362
QY 301 gggcgcgcgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 360
Db 363 gggcgcgcgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 422
QY 361 acagggacgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 396
Db 423 acagggacgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 458

RESULT 5

AA199683
ID AA199683 standard; DNA; 4403765 BP.

AC AA199683;

DT 15-JAN-2002 (first entry)

DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.

KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
variation; epidemiology; patient treatment; epidemic monitoring; ds.

OS Mycobacterium tuberculosis.

PN US6294328-B1. *M. tuberculosis*

PD 25-SEP-2001.

PE 24-JUN-1998; 98US-0103840.

PR 24-JUN-1998; 98US-0103840.

PA (GENO-) INST GENOMIC RES.

PI Fleischmann RD, White OR, Fraser CM, Venter JC;

DR WPI; 2001-647261/74.

PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
determining the nucleotide sequence of the strain at positions in the

PT genome corresponding to positions where M. tuberculosis strains CDC
1551 and H37Rv differ

PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.

CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?DocID=6294328B1.

CC Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;

Query Match 100.0%; Score 396; DB 22; Length 4403765;

Best Local Similarity 100.0%; Pred. No. 4e-85; Mismatches 0; Indels 0; Gaps 0;

QY 1 acggcgcgctccgataactcagctgtcccaagggtggcaggatctcgcatccgac 60
Db 151984 acggcgcgctccgataactcagctgtcccaagggtggcaggatctcgcatccgac 152043
QY 61 gggcgcgcgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 120
Db 152044 gggcgcgcgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 152103
QY 121 atcggcgcgcgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 180
Db 152104 atcggcgcgcgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 152163
QY 181 gtccacgcgtgtgctggagcgctccgcgcgaagtcttcgcatccacggcgagctg 240
Db 152164 gtccacgcgtgtgctggagcgctccgcgcgaagtcttcgcatccacggcgagctg 152223
QY 241 atcacgcggttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 300
Db 152224 atcacgcggttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 152283
QY 301 gggcgcgcgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 360
Db 152284 gggcgcgcgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 152343
QY 361 acagggacgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 396
Db 152344 acagggacgttcgacgcgcctccgataactcgcgcacgcgcgacgttacc 152379

RESULT 6

AA199682
ID AA199682 standard; DNA; 4411529 BP.

AC AA199682;

DT 15-JAN-2002 (first entry)

DE Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 1.

KM Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
variation; epidemiology; patient treatment; epidemic monitoring; ds.

OS Mycobacterium tuberculosis.

PN US6294328-B1. *M.*

PD 25-SEP-2001.

PE 24-JUN-1998; 98US-0103840.

PR 24-JUN-1998; 98US-0103840.

PA (GENO-) INST GENOMIC RES.

PI Fleischmann RD, White OR, Fraser CM, Venter JC;

DR WPI; 2001-647261/74.

PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
determining the nucleotide sequence of the strain at positions in the

PT genome corresponding to positions where M. tuberculosis strains CDC
1551 and H37Rv differ

PS Claim 3; SEQ ID NO 1; 3pp + Sequence Listing; English.

CC The invention relates to evaluating strain variation within and between
CC different populations of the tuberculosis bacterial pathogen,
CC Mycobacterium tuberculosis or related Mycobacterium by determining the
CC nucleotide sequence of the first strain at positions in the complete
CC sequence of the genome that correspond to positions that differ in the
CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA199683) and
CC H37Rv (AA199682). The method is useful for evaluating strain variation of

KW skin testing; M.tuberculosis; ss.
 XX Mycobacterium tuberculosis.
 OS
 FH Key Location/Qualifiers
 FT CDS 11..409 /*tag= a
 FT /product= Antigen_TbRa12
 PN MO9709429-A2.
 XX
 PD 13-MAR-1997.
 PF 30-APR-1996; 96WO-US14675.
 XX
 PR 12-JUL-1996; 96US-0680573.
 PR 01-SEP-1995; 95US-0523435.
 PR 22-SEP-1995; 95US-0532136.
 PR 22-MAR-1996; 96US-0620280.
 PR 05-JUN-1996; 96US-0658800.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAW;
 PI Twardzik DR, Vedvick TH;
 XX
 DR WPI; 1997-192904/17.
 DR P-PSDB; AAW32354.
 XX
 PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens
 PT - useful for diagnosis of M. tuberculosis infection
 XX
 PS Claim 3; Page 49-50; 190pp; English.
 XX
 CC A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
 CC its variant differing only in conservative substitutions and/or
 CC modifications). The present sequence encodes a specifically claimed
 CC M.tuberculosis antigen, TbRa12. The immunogenic polypeptide can be
 CC used to diagnose M.tuberculosis infection by forming complexes with
 CC specific antibodies in the sample. Fragments of DNA encoding the
 CC immunogenic polypeptide can be used as diagnostic primers or probes
 CC and agents that bind to the antigen, especially monoclonal antibodies
 CC or equivalent polyclonal antibodies, are also used for diagnosis.
 CC
 XX
 SQ Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;
 SQ

Query Match 99.6%; Score 394.4; DB 18; Length 447;
 Best local Similarity 99.7%; Pred. No. 4.7e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggcgcggtccgatactccagctgtcccgaggtggtgagcagagattccattccgac 60
 DB 11 acggcgcggtccgatactccagctgtcccgaggtggtgagcagagattccattccgac 70
 QY 61 gggcagcgatgagc 120
 DB 71 gggcagcgatgagc 130
 QY 121 atggggctaccggctctccctcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
 DB 131 atggggctaccggctctccctcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 190
 QY 181 gtccaaagcgt 240
 DB 191 gtccaaagcgt 250
 QY 241 ataccggcgatgagc 300
 DB 251 ataccggcgatgagc 310
 QY 301 gggcattcattccggtgagctatctcgtgtgacctgtgcaaaccaagtcgggcgagcgcgt 360

DB 311 gggcattcattccggtgagctatctcgtgtgacctgtgcaaaccaagtcgggcgagcgcgt 370
 QY 361 acagggagcgtgacattgtgacgagcagcccgccgac 396
 DB 371 acagggagcgtgacattgtgacgagcagcccgccgac 406

RESULT 9
 AAV64450
 ID AAV64450 standard; DNA; 447 BP.
 XX
 AC AAV64450;
 XX
 DT 27-JAN-1999 (first entry)
 XX
 DE M. tuberculosis immunogenic polypeptide TbRa12 DNA.
 XX
 KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis; ss.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9816646-A2.
 PD 23-APR-1998.
 PF 07-OCT-1997; 97WO-US18293.
 XX
 PR 13-MAR-1997; 97US-0818112.
 PR 11-OCT-1996; 96US-0730510.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Houghton R, Lodes MJ;
 PI Reed SG, Skeiky YAW, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1998-261042/23.
 DR P-PSDB; AAW81657.
 XX
 PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used
 PT to develop products for the detection of M. tuberculosis infection
 PT and for diagnosis, treatment and prevention of tuberculosis
 XX
 PS Claim 3; Page 62-63; 230pp; English.
 XX
 CC This sequence encodes an immunogenic portion of a soluble Mycobacterium
 CC tuberculosis (MT) antigen which can be used in a method for inducing
 CC protective immunity against tuberculosis (TB). This sequence can be
 CC formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis.
 CC
 XX
 SQ Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;
 SQ

Query Match 99.6%; Score 394.4; DB 19; Length 447;
 Best local Similarity 99.7%; Pred. No. 4.7e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggcgcggtccgatactccagctgtcccgaggtggtgagcagagattccattccgac 60
 DB 11 acggcgcggtccgatactccagctgtcccgaggtggtgagcagagattccattccgac 70
 QY 61 gggcagcgatgagc 120
 DB 71 gggcagcgatgagc 130
 QY 121 atggggctaccggctctccctcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 180
 DB 131 atggggctaccggctctccctcggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 190
 QY 181 gtccaaagcgt 240

CC part of a Mycobacterium tuberculosis antigen (Ag). Also described
 CC are vaccines and fusion protein containing M. tuberculosis Ag's.
 CC M. tuberculosis Ag's, DNAs encoding them, derived fusion proteins and
 CC other polypeptides fragments, can be used in pharmaceutical compositions
 CC or vaccines to generate a protective or therapeutic immune response to
 CC M. tuberculosis and as reagents in skin tests for diagnosis of
 CC tuberculosis. Ag can induce proliferation of, or cytokine secretion
 CC by, T, B or natural killer cells and/or macrophages in
 CC tuberculosis-immune subjects. AA219249 to AA219460 and AA219083 to
 CC AA219225 are used in the exemplification of the present invention.

XX Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;

Query Match 99.6%; Score 394.4; DB 20; Length 447;
 Best Local Similarity 99.7%; Pred. No. 4.7e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggccgctcgcataactccagctgtcccaagggtggcaggatcgccatccgac 60
 DB 11 acggccgctcgcataactccagctgtcccaagggtggcaggatcgccatccgac 70
 QY 61 ggccagcagatggcagatccgcggccagatcgatcggtggggggtcaccacgctcat 120
 DB 71 ggccagcagatggcagatccgcggccagatcgatcggtggggggtcaccacgctcat 130
 QY 121 atcgggcctacgcgctctcctcggtgtgtgttcgaaacaaacggcagcgcgacga 180
 DB 131 atcgggcctacgcgctctcctcggtgtgtgttcgaaacaaacggcagcgcgacga 190
 QY 181 gtccaaagcgtgtgtcggagcgtcccgcggaagltcgcgcatccacggcgacgtg 240
 DB 191 gtccaaagcgtgtgtcggagcgtcccgcggaagltcgcgcatccacggcgacgtg 250
 QY 241 ataccggcggtgcagcggtcgcgcatccagtcggccacggcgatggcgagcgcttaac 300
 DB 251 ataccggcggtgcagcggtcgcgcatccagtcggccacggcgatggcgagcgcttaac 310
 QY 301 ggccatcatcccggtgacgtatcgtgtgacctggcacaacgaatcgggcgacgct 360
 DB 311 ggccatcatcccggtgacgtatcgtgtgacctggcacaacgaatcgggcgacgct 370
 QY 361 acagggaacgtgacatggcgagggagcccgccg 396
 DB 371 acagggaacgtgacatggcgagggagcccgccg 406

RESULT 12

AA219040
 ID AA219040 standard; DNA; 447 BP.

XX
 AC AA219040;

XX
 DT 05-NOV-1999 (first entry)

DE M. tuberculosis recombinant antigen DNA encoding TbrA12.

XX Antigen; diagnosis; detection; infection; antibody; immunisation;

KW vaccine; immunity; ss.

XX Mycobacterium tuberculosis.

XX MO9942118-A2.

XX PD 26-AUG-1999.

XX PF 17-FEB-1999; 99WO-US03265.

XX PR 05-MAY-1998; 98US-0072596.

XX PR 18-FEB-1998; 98US-0024753.

XX PA (CORR-) CORIXA CORP.

PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R,
 PI Lodes MJ, Reed SG, Skelky YAM, Twardzik DR, Vedralik TS;
 XX WP1: 1999-527416/44.
 DR P-SDB; AAY38959.

XX New polypeptide comprising antigenic portions of M. tuberculosis
 PS Claim 3; Page 103; 323pp; English.

CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.

XX Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;

Query Match 99.6%; Score 394.4; DB 20; Length 447;
 Best Local Similarity 99.7%; Pred. No. 4.7e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 acggccgctcgcataactccagctgtcccaagggtggcaggatcgccatccgac 60
 DB 11 acggccgctcgcataactccagctgtcccaagggtggcaggatcgccatccgac 70
 QY 61 ggccagcagatggcagatccgcggccagatcgatcggtggggggtcaccacgctcat 120
 DB 71 ggccagcagatggcagatccgcggccagatcgatcggtggggggtcaccacgctcat 130
 QY 121 atcgggcctacgcgctctcctcggtgtgtgttcgaaacaaacggcagcgcgacga 180
 DB 131 atcgggcctacgcgctctcctcggtgtgtgttcgaaacaaacggcagcgcgacga 190
 QY 181 gtccaaagcgtgtgtcggagcgtcccgcggaagltcgcgcatccacggcgacgtg 240
 DB 191 gtccaaagcgtgtgtcggagcgtcccgcggaagltcgcgcatccacggcgacgtg 250
 QY 241 ataccggcggtgcagcggtcgcgcatccagtcggccacggcgatggcgagcgcttaac 300
 DB 251 ataccggcggtgcagcggtcgcgcatccagtcggccacggcgatggcgagcgcttaac 310
 QY 301 ggccatcatcccggtgacgtatcgtgtgacctggcacaacgaatcgggcgacgct 360
 DB 311 ggccatcatcccggtgacgtatcgtgtgacctggcacaacgaatcgggcgacgct 370
 QY 361 acagggaacgtgacatggcgagggagcccgccg 396
 DB 371 acagggaacgtgacatggcgagggagcccgccg 406

RESULT 13

AA503780
 ID AA503780 standard; DNA; 447 BP.

XX
 AC AA503780;

XX
 DT 29-AUG-2001 (first entry)

DE M. tuberculosis DNA encoding a partial antigen TbrA12.

XX TbrA12; antigen; vaccine; tuberculosis; AIDS;

KW acquired immunodeficiency disease; ss.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers

FT 11. 409 /tag= a

FT CDS /product= "TbrA12"


```

FT FT /partial
XX XX /note="No start codon"
PN WO200124820-A1.
PD 12-APR-2001.
XX 10-OCT-2000; 2000WO-US28095.
XX 07-OCT-1999; 99US-0158338.
XX 07-OCT-1999; 99US-0158425.
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX MPI; 2001-290576/30.
XX P-PSDB; AAU01889.
XX
XX Vaccinating against Mycobacteria infections in mammals using fusion
XX proteins comprising combinations of heterologous antigens
XX
XX Example 2; Page 162; 168pp; English.
XX
XX The sequence encodes Mycobacterium tuberculosis TbrA12,
XX an M. tuberculosis antigen. Compositions comprising at least
XX 2 heterologous antigens, as a fusion protein, and vectors expressing the
XX fusion proteins are used as vaccines to prophylactically immunise
XX mammals (especially humans) against infection by Mycobacteria. The
XX compositions contain at least 2 heterologous antigens that increase the
XX serological sensitivity of individuals infected with tuberculosis, a
XX disease frequently affecting patients with acquired immunodeficiency
XX disease, AIDS.
XX
XX Sequence 447 BP; 79 A; 146 C; 149 G; 72 T; 1 other;
SQ

```

Query Match 99.6%; Score 394.4; DB 22; Length 447;
 Best Local Similarity 99.7%; Pred. No. 4,7e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

OY 1 acggcgcgtccgaatactccagctgtccaggggtggcaggatccgcatccgcatc 60
DB 11 acggcgcgtccgaatactccagctgtccaggggtggcaggatccgcatccgcatc 70
OY 61 ggccagcgtatggtcgatcgcgccagatccgcatcggtgggggtccaccacggttcat 120
DB 71 ggccagcgtatggtcgatcgcgccagatccgcatcggtgggggtccaccacggttcat 130
OY 121 atcgggctaccgctctcgtcgttgggtgtgtgtcgacaacagcgacggcgacga 180
DB 131 atcgggctaccgctctcgtcgttgggtgtgtgtcgacaacagcgacggcgacga 190
OY 181 gtccaaacgtgtgtcgatcggtcgccagagtcctcgcatctccaccggtcgagt 240
DB 191 gtccaaacgtgtgtcgatcggtcgccagagtcctcgcatctccaccggtcgagt 250
OY 241 ataccggcggtcgacggcggtccgataactccggcaccggcggatggcgacgcttaac 300
DB 251 ataccggcggtcgacggcggtccgataactccggcaccggcggatggcgacgcttaac 310
OY 301 ggacatcatcccggtgacgtatcatcgtgtgacctggtgcaaacaaagtcggcgacggt 360
DB 311 ggacatcatcccggtgacgtatcatcgtgtgacctggtgcaaacaaagtcggcgacggt 370
OY 361 acagggaacgtgacatgtggccaggagaccggcgcc 396
DB 371 acagggaacgtgacatgtggccaggagaccggcgcc 406

```

RESULT 14
 AAS03793
 ID AAS03793 standard; DNA; 1629 BP.

```

XX AC AAS03793;
XX XX
XX 29-AUG-2001 (first entry)
XX XX
XX M. tuberculosis DNA encoding antigen HTCC#1 fusion protein #4.
XX XX
XX TbrA12-HTCC#1; antigen; vaccine; tuberculosis;
XX KW AIDS; acquired immunodeficiency disease; His tag; ds.
XX XX
XX Mycobacterium tuberculosis.
XX OS Synthetic.
XX
XX Key location/Qualifiers
XX CDS 1..1629
XX
XX /*tag= a "TbrA12-HTCC#1"
XX /*product= (pos:1621..1623,aa:xaa)
XX /*transl_except= (pos:1621..1623,aa:xaa)
XX /*note= "Xaa" in frame STOP codon"
XX /*partial
XX /*note= "No start or stop codon. Although the sequence
XX does contain an in frame stop codon, 2 further amino
XX acids are shown in figure 8 as being encoded by the
XX present sequence, without a further stop codon"
XX 25..426
XX /*tag= b
XX /*note= "Region derived from TbrA12"
XX 427..444
XX /*tag= c
XX /*note= "Region derived from Thrombin"
XX 445..1629
XX /*tag= d
XX /*note= "Region derived from HTCC#1"
XX
XX WO200124820-A1.
XX
XX 12-APR-2001.
XX
XX 10-OCT-2000; 2000WO-US28095.
XX XX
XX 07-OCT-1999; 99US-0158338.
XX XX
XX 07-OCT-1999; 99US-0158425.
XX XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX MPI; 2001-290576/30.
XX
XX Vaccinating against Mycobacteria infections in mammals using fusion
XX proteins comprising combinations of heterologous antigens
XX
XX Example 2; Fig 8; 168pp; English.
XX
XX The sequence encodes Mycobacterium tuberculosis fusion protein,
XX TbrA12-HTCC#1 and includes a His tag at the N-terminus to aid
XX purification. Compositions comprising at least 2 heterologous antigens,
XX as a fusion protein, and vectors expressing the fusion proteins are
XX used as vaccines to prophylactically immunise mammals (especially
XX humans) against infection by Mycobacteria. The compositions
XX contain at least 2 heterologous antigens that increase the serological
XX sensitivity of individuals infected with tuberculosis, a disease
XX frequently affecting patients with acquired immunodeficiency disease,
XX AIDS.
XX
XX Sequence 1629 BP; 298 A; 500 C; 535 G; 296 T; 0 other;
SQ

```

Query Match 99.6%; Score 394.4; DB 22; Length 1629;
 Best Local Similarity 99.7%; Pred. No. 5.2e-85;
 Matches 395; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 acggcgcgtccgaatactccagctgtccaggggtggcaggatccgcatccgcatc 60

```

Db      25  acgcccgcgtccgataactccagctgtccaccaggtgtggaaggaattccatccgatac 84
QY      61  gggcagcgatggtcgatccgagccagatccgagctggtgggggttcaaccacgttcat 120
Db      85  gggcagcgatggtcgatccgagccagatccgagctggtgggggttcaaccacgttcat 144
QY     121  atcgagctacacgcttccctcgcttgggtgttctcgacacaaacgaagcgagcaga 180
Db     145  atcgagctacacgcttccctcgcttgggtgttctcgacacaaacgaagcgagcaga 204
QY     181  gtccacacgctgtgtcgagcgctccgagcgagatctcgatccacccgagcagt 240
Db     205  gtccacacgctgtgtcgagcgctccgagcgagatctcgatccacccgagcagt 264
QY     241  ataccgaggtgagcagcgctccgataactccgacccgagatgagcagcgttaac 300
Db     265  ataccgaggtgagcagcgctccgataactccgacccgagatgagcagcgttaac 324
QY     301  gggcatalatcccggtgagctcgtgacgttgcaacaaacgaagtcggcgagcagcgt 360
Db     325  gggcatalatcccggtgagctcgtgacgttgcaacaaacgaagtcggcgagcagcgt 384
QY     361  acagggaaactgacatggtcgagcgagaccccgagc 396
Db     385  acagggaaactgacatggtcgagcgagaccccgagc 420

RESULT  15
AAT91477
ID      AAT91477 standard; DNA; 1872 BP.
AC      AAT91477;
XX
DT      07-JAN-1998 (first entry)
XX
DE      Mycobacterium tuberculosis antigen T8Ra35 encoding DNA.
XX
KW      Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX      skin testing; M.tuberculosis; ss.
XX
OS      Mycobacterium tuberculosis.
XX
FH      Key      Location/Qualifiers
FT      CDS      89..1156
FT              /tag="a
FT              /product=" Antigen_T8Ra35
XX
XX      WO9709428-A2.
XX
XX      PD      13-MAR-1997.
XX
XX      PE      30-AUG-1996; 96WO-US14674.
XX
XX      PR      12-JUL-1996; 96US-0680574.
XX      PR      01-SEP-1995; 95US-0523436.
XX      PR      22-SEP-1995; 95US-0533634.
XX      PR      22-MAR-1996; 96US-0620874.
XX      PR      05-JUN-1996; 96US-0659683.
XX
XX      PA      (CORI-) CORIXA CORP.
XX
XX      PI      Campos-neto A, Dillon DC, Houghton R, Reed SG, Skeiky YAM;
XX      PI      Twardzik DR, Vedyck TH;
XX
XX      DR      WPI; 1997-192903/17.
XX      DR      P-PSDB; AAW32435.
XX
XX      PT      New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
XX      PT      useful in vaccines for prevention or treatment of tuberculosis, also
XX      PT      for diagnosis
XX
XX      PS      Claim 3; Page 63-65; 168pp; English.

```

```

XX
CC      A new immunogenic polypeptide has been developed comprising an
CC      immunogenic part of a soluble Mycobacterium tuberculosis antigen (or
CC      its variant differing only in conservative substitutions and/or
CC      modifications). The present sequence encodes a specifically claimed
CC      M.tuberculosis antigen, T8Ra35. The immunogenic protein, and fusion
CC      proteins containing one or more of the proteins or one of the proteins
CC      plus ESAT-6, are useful in vaccines, preferably when formulated with a
CC      non-specific adjuvant, to induce an immune response against
CC      M.tuberculosis (for treatment or prevention).
XX
SO      Sequence 1872 BP; 318 A; 616 C; 604 G; 332 T; 2 other:

Query Match      98.8%; Score 391.2; DB 18; Length 1872;
Best Local Similarity 99.2%; Pred. No. 3,1e-84;
Matches 393; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      1  acgcccgcgtccgataactccagctgtccaccaggtgtggaaggaattccatccgatac 60
Db      758  acgcccgcgtccgataactccagctgtccaccaggtgtggaaggaattccatccgatac 817
QY      61  gggcagcgatggtcgatccgagccagatccgagctggtgggggttcaaccacgttcat 120
Db      818  gggcagcgatggtcgatccgagccagatccgagctggtgggggttcaaccacgttcat 877
QY      121  atcgagctacacgcttccctcgcttgggtgttctcgacacaaacgaagcgagcaga 180
Db      878  atcgagctacacgcttccctcgcttgggtgttctcgacacaaacgaagcgagcaga 937
QY      181  gtccacacgctgtgtcgagcgctccgagcgagatctcgatccacccgagcagt 240
Db      938  gtccacacgctgtgtcgagcgctccgagcgagatctcgatccacccgagcagt 997
QY      241  ataccgaggtgagcagcgctccgataactccgacccgagatgagcagcgttaac 300
Db      998  ataccgaggtgagcagcgctccgataactccgacccgagatgagcagcgttaac 1057
QY      301  gggcatalatcccggtgagctcgtgacgttgcaacaaacgaagtcggcgagcagcgt 360
Db     1058  gggcatalatcccggtgagctcgtgacgttgcaacaaacgaagtcggcgagcagcgt 1117
QY      361  acagggaaactgacatggtcgagcgagaccccgagc 396
Db     1118  acagggaaactgacatggtcgagcgagaccccgagc 1153

```

Search completed: August 6, 2002, 11:39:56
 Job time: 6866 sec

DR WPI, 1999-601610/51.
DR N-PSDB; AA220206.

PT New fusion proteins useful for diagnosis, prevention and treatment of
PT tuberculosis -

PS Claim 1; Fig 13A-B; 83pp; English.

XX This sequence represents a recombinant Mycobacterium tuberculosis
CC bi-antigen fusion protein, termed MtB24, composed of the antigens
CC Ra12 and DDPd. The fusion protein is expressed in host cells
CC using a vector carrying a polynucleotide (see AA220206) comprising
CC the coding sequences for the 2 antigens. The invention provides
CC fusion proteins (see AA22059-71) containing at least 2 M.
CC tuberculosis antigens. The new fusion proteins and polynucleotides
CC encoding them are useful as vaccines for preventing tuberculosis
CC (claimed), for diagnosis (via in vitro assays or intradermal skin
CC tests for detection of anti-M. tuberculosis antibodies), monitoring
CC of disease progression, and treatment of tuberculosis. They are
CC more effective immunogens than mixtures of the individual protein
CC components.

XX Sequence 231 AA;

Query Match 100.0%; Score 675; DB 20; Length 231;
Best Local Similarity 100.0%; Pred. No. 4.8e-59;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAADNPFOLSGGCGFAIPIGOMAIAGQIRSGGSPYHIGPTAPLGGVYDNNNGAR 60
Db 9 taasdnfqlsggggfgfaiipigamalaqfrrsggsprthipgtalilgyvndngngar 68

QY 61 VQRYVGSAPASISIGTGVITAVDGPINSATAMADALNGHHPGVISVTWQTSGGTR 120
Db 69 vqrvvgsapasaigstgvtavdgapinsatamadalinghpgvsvtwqtsggtr 128

QY 121 TGNVTLAEGPPA 132
Db 129 tgnvltlaegppa 140

RESULT 2

AAV05000
ID AAV05000 standard; Protein; 355 AA.

AC AAV05000;

DT 06-JUL-1999 (first entry)

XX Mycobacterium species protein sequence 50D.

KM Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
KM hybridisation; detection; vaccine; immunisation; infection.

OS Mycobacterium sp.

PN WO909186-A2.

PD 25-FEB-1999.

PF 14-AUG-1998; 98WO-FR01813.

PR 11-SEP-1997; 97FR-0011325.

PR 14-AUG-1997; 97FR-0010404.

PA (INSP) INSR PASTEUR.

PI Gicquel B, Lhm EM, Pelicic V, Portnoi D, Goguet de la Salmoniere Y;

DR WPI, 1999-181045/15.
DR N-PSDB; AA34251.

XX Mycobacterial DNA vectors containing reporter constructs - for
PT identifying coding or promoter sequences involved in
PT infection-associated protein expression

PS Claim 32; Fig 50D; 309pp; French.

XX Sequences AAV04742-Y05000 and AAV07201-Y07204 represent secreted
CC proteins from various Mycobacterium species microorganisms. The
CC encoding nucleotide sequences can be used as primers and probes for
CC methods for detecting and identifying mycobacteria, especially belonging
CC to the M. tuberculosis complex. The encoded proteins can be used in
CC vaccines for immunisation against a bacterial or viral infection.

XX Sequence 355 AA;

Query Match 100.0%; Score 675; DB 20; Length 355;
Best Local Similarity 100.0%; Pred. No. 8.3e-59;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAADNPFOLSGGCGFAIPIGOMAIAGQIRSGGSPYHIGPTAPLGGVYDNNNGAR 60
Db 224 taasdnfqlsggggfgfaiipigamalaqfrrsggsprthipgtalilgyvndngngar 283

QY 61 VQRYVGSAPASISIGTGVITAVDGPINSATAMADALNGHHPGVISVTWQTSGGTR 120
Db 284 vqrvvgsapasaigstgvtavdgapinsatamadalinghpgvsvtwqtsggtr 343

QY 121 TGNVTLAEGPPA 132
Db 344 tgnvltlaegppa 355

RESULT 3

AAG81110
ID AAG81110 standard; Protein; 355 AA.

AC AAG81110;

DT 04-SEP-2001 (first entry)

XX Mycobacterium tuberculosis potential drug target protein SEQ ID 161.

KM Drug target; growth; organism viability; characterisation.

OS Mycobacterium tuberculosis.

PN WO200135317-A1.

PD 17-MAY-2001.

PF 13-NOV-2000; 2000MO-US31152.

PR 12-NOV-1999; 99US-0165086.

PR 12-NOV-1999; 99US-0165124.

PR 01-FEB-2000; 2000US-0179531.

PA (REGC) UNIV CALIFORNIA.

PI Eisenberg D, Rotstein SH, Marcotte EM;

DR WPI, 2001-329193/34.

DR N-PSDB; AAH51961.

PT Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the
PT sequences -

PS Disclosure; Page 157; 207pp; English.

XX This invention relates to a method for identifying a nucleotide or
CC

CC polypeptide sequence that may be a drug target, or essential for growth
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
 CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analysing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism.

XX Sequence 355 AA:

Query Match 100.0%; Score 675; DB 22; Length 355;
 Best Local Similarity 100.0%; Pred. No. 8.3e-59;
 Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAADNFOLSGGGGFAIRPIGAMATAGQIRSGGSPYHIGPTAFILGVDNNGNGAR 60
 DB 224 taasdnfqlsggggfaipigemalaqtrsggsptvhiqptafilgvdnngngar 283
 QY 61 VORVGSAPAAASLGISTGVDITAVDGAIPINSATAMADALNGHHPGVISVTWQKSGGTR 120
 DB 284 vqrvvgsapaaslgistgvtavdgaplnsatamadalnghhpgdvlsvtwtksqgtr 343
 QY 121 TGNVTLAEGPPA 132
 DB 344 tgnvltlaegppa 355

RESULT 4

AAU04830
 ID AAU04830 standard; Protein; 379 AA.

XX AAU04830;

DT 06-JUL-1999 (first entry)

XX Mycobacterium species protein sequence 50F.

XX Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.

XX Mycobacterium sp.

OS W09909186-A2.

XX W09909186-A2.

XX 25-FEB-1999.

XX 14-AUG-1998; 98WO-FR01813.

XX 11-SEP-1997; 97FR-0011325.

XX 14-AUG-1997; 97FR-0010404.

XX (INSP) INST PASTEUR.

XX Glacquel B, Lim EM, Pellicic V, Portnoi D, Goguet de la Salmoniere Y;

XX Guigueno A;

XX WPI; 1999-181045/15.

XX N-PSDB; AAX34252.

XX Mycobacterial DNA vectors containing reporter constructs - for

XX identifying coding or promoter sequences involved in

XX infection-associated protein expression

XX Claim 32; Fig 50F; 309pp; French.

XX Sequences AAU04742-Y05000 and AAU07201-Y07204 represent secreted

XX proteins from various Mycobacterium species microorganisms. The

CC encoding nucleotide sequences can be used as primers and probes for
 CC methods for detecting and identifying mycobacteria, especially belonging
 CC to the M. tuberculosis complex. The encoded proteins can be used in
 CC vaccines for immunisation against a bacterial or viral infection.

XX Sequence 379 AA:

Query Match 100.0%; Score 675; DB 20; Length 379;
 Best Local Similarity 100.0%; Pred. No. 9e-59;
 Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAADNFOLSGGGGFAIRPIGAMATAGQIRSGGSPYHIGPTAFILGVDNNGNGAR 60
 DB 248 taasdnfqlsggggfaipigemalaqtrsggsptvhiqptafilgvdnngngar 307
 QY 61 VORVGSAPAAASLGISTGVDITAVDGAIPINSATAMADALNGHHPGVISVTWQKSGGTR 120
 DB 308 vqrvvgsapaaslgistgvtavdgaplnsatamadalnghhpgdvlsvtwtksqgtr 367
 QY 121 TGNVTLAEGPPA 132
 DB 368 tgnvltlaegppa 379

RESULT 5

AAU01905
 ID AAU01905 standard; Protein; 543 AA.

XX AAU01905;

DT 29-AUG-2001 (first entry)

XX M. tuberculosis antigen H7CC#1 fusion protein #4.

XX TBRA12-H7CC#1; antigen; vaccine; tuberculosis;

XX AIDS; acquired immunodeficiency disease; His Tag.

XX Mycobacterium tuberculosis.

OS Synthetic.

XX Key

XX Binding-site

XX Location/Qualifiers

XX 3..8

XX /label= "Histidine-tag

XX /note= "Nickel chelating region used to aid

XX purification of the protein"

XX Msc-difference 541

XX /label= OTHER

XX /note= "In frame STOP codon"

XX W0200124820-A1.

XX 12-Apr-2001.

XX 10-OCT-2000; 2000WO-US28095.

XX 07-OCT-1999; 99US-0158338.

XX 07-OCT-1999; 99US-0158425.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;

XX WPI; 2001-290576/30.

XX Vaccinating against Mycobacteria infections in mammals using fusion

XX proteins comprising combinations of heterologous antigens -

XX Claim 17; Fig 8; 168pp; English.

XX The sequence represents Mycobacterium tuberculosis fusion protein,

XX TBRA12-H7CC#1 and includes a His tag at the N-terminus to aid

XX purification. Compositions comprising at least 2 heterologous

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 99.3%; Score 670; DB 20; Length 132;
 Best Local Similarity 99.2%; Pred. No. 7.4e-59;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAASDNFOLSGGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 taasdnfqlsggggqfaipigqamalaqqlrsggspvthigpqlafglgvvdnngngar 60

OY 61 VORVGSAPASAGISTGDIYTAVDGAPINSATAMADALNGHHHPGDVISTWQTKSGCTR 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 vqrvvgsapaasagistgdiytavdgapinsatamadalnghpgdvisvwnqtksgtr 120

OY 121 TGNVTIAEGPPA 132
 ||||||||||||
 Db 121 tgnvtlaegppa 132

RESULT 12
 AAY38959
 ID AAY38959 standard; Protein; 132 AA.
 XX
 AC AAY38959;
 XX
 DT 05-NOV-1999 (first entry)
 XX
 DE M. tuberculosis recombinant antigen protein TbrA12.
 XX
 KM Antigen; diagnosis; detection; infection; antibody; immunisation;
 KM vaccine; immunity.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W09942118-A2.
 PD 26-AUG-1999.
 PF 17-FEB-1999; 99MO-US03265.
 XX
 PR 05-MAY-1998; 98US-0072596.
 PR 18-FEB-1998; 98US-0024753.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Campos-Neto A, Dillon DC, Hendrickson RC, Houghton R;
 PI Lodes MJ, Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;
 XX
 DR WPI; 1999-527416/44.
 DR N-PSDB; AA219040.
 XX
 PT New polypeptide comprising antigenic portions of M. tuberculosis
 XX
 PS Example 3; Page 138; 323pp; English.
 XX
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis infection in a
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against
 CC M. tuberculosis infection. The new detection methods are needed as
 CC current vaccination strategies do not provide 100% immunity.
 CC
 XX
 SO Sequence 132 AA;

Query Match 99.3%; Score 670; DB 20; Length 132;
 Best Local Similarity 99.2%; Pred. No. 7.4e-59;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAASDNFOLSGGGGFAIPIGQAMAIAQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 taasdnfqlsggggqfaipigqamalaqqlrsggspvthigpqlafglgvvdnngngar 60

OY 61 VORVGSAPASAGISTGDIYTAVDGAPINSATAMADALNGHHHPGDVISTWQTKSGCTR 120
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 61 vqrvvgsapaasagistgdiytavdgapinsatamadalnghpgdvisvwnqtksgtr 120

OY 121 TGNVTIAEGPPA 132
 ||||||||||||
 Db 121 tgnvtlaegppa 132

RESULT 13
 AAU69898
 ID AAU69898 standard; Protein; 132 AA.
 XX
 AC AAU69898;
 XX
 DT 30-JAN-2002 (first entry)
 XX
 DE Mycobacterium tuberculosis antigen Ra12.
 XX
 KM Prostate cancer; cytostatic; immunostimulant; tumour; immunogen;
 KM fusion protein; Ra12 antigen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN W0200173032-A2.
 PD 04-OCT-2001.
 PF 27-MAR-2001; 2001WO-US09919.
 XX
 PR 27-MAR-2000; 2000US-0536857.
 PR 09-MAY-2000; 2000US-0568100.
 PR 12-MAY-2000; 2000US-0570737.
 PR 13-JUN-2000; 2000US-0593793.
 PR 27-JUN-2000; 2000US-0605783.
 PR 10-AUG-2000; 2000US-0636215.
 PR 29-AUG-2000; 2000US-0651236.
 PR 06-SEP-2000; 2000US-0657279.
 PR 02-OCT-2000; 2000US-0679426.
 PR 10-OCT-2000; 2000US-0685166.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Reller MM, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAM, Hepler WT, Henderson RA;
 XX
 DR WPI; 2001-639232/73.
 XX
 PT New human prostate-specific polypeptides and polynucleotides useful for
 PT the diagnosis and treatment of cancer, especially prostate cancer -
 XX
 PS Example 17; Page 531-532; 579pp; English.
 XX
 CC The invention relates to isolated prostate-specific
 CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
 CC antibodies raised against the polypeptides (or antigenic epitopes
 CC derived from them) and antigen presenting cells expressing the
 CC polypeptides. The antibodies are useful for detecting the presence of
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
 CC the antigen-presenting cells are useful for stimulating and/or expanding
 CC T cells specific for a tumour protein, and for inhibiting the development
 CC of cancer especially prostate cancer. Compositions comprising the
 CC polynucleotide and/or polypeptide are useful for stimulating an immune
 CC response, and for treating cancer. The oligonucleotide is useful for
 CC detecting cancer. The present sequence is M. tuberculosis Ra12
 CC antigen (or fragment) used to make a fusion protein comprising a
 CC prostate specific polypeptide of the invention.
 CC
 XX
 SO Sequence 132 AA;

Query Match 99.3%; Score 670; DB 22; Length 132;
 Best Local Similarity 99.2%; Pred. No. 7.4e-59;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGGFAIPICGAMATAAGQIRSGGSPVTHIGTATLGLGVNDNNGAR 60
 |||||||
 Db 1 taasdnfqlsggggfaipigamataagqlrsggspvthlglglgvndnngar 60
 |||||||
 QY 61 VORVVGSAFPAASLIGSTGDIITAVDGPINSATAMADALNGHHPGDVISTWTKSGGTR 120
 |||||||
 Db 61 vgrvvgsapaasligstgdiitavdgaplnsatamadalnghpgdvisvnmwtksggtr 120
 |||||||
 QY 121 TGNVTLAEGPPA 132
 |||||||
 Db 121 tgnvtlaegppa 132

RESULT 14
 ID AAU69906 standard; protein; 132 AA.

AC AAU69906;
 30-JAN-2002 (first entry)

DE Mycobacterium tuberculosis antigen Ra12.
 KM Prostate cancer; cytostatic; immunostimulant; tumour; immunogen;
 fusion protein; Ra12 antigen.

OS Mycobacterium tuberculosis.
 PN WO200173032-A2.
 PD 04-OCT-2001.

PF 27-MAR-2001; 2001MO-US09919.
 XX 27-MAR-2000; 2000US-0536857.
 PR 09-MAY-2000; 2000US-0568100.
 PR 12-MAY-2000; 2000US-0570737.
 PR 13-JUN-2000; 2000US-0593793.
 PR 27-JUN-2000; 2000US-0605783.
 PR 10-AUG-2000; 2000US-0636215.
 PR 29-AUG-2000; 2000US-0651236.
 PR 06-SEP-2000; 2000US-0657279.
 PR 02-OCT-2000; 2000US-0679426.
 PR 10-OCT-2000; 2000US-0685166.

PA (CORI-) CORIXA CORP.
 XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Jiang Y, Kalos MD;
 PI Fanger GR, Retter MW, Stolk JA, Day CH, Vedvick TS, Carter D;
 PI Li SX, Wang A, Skeiky YAW, Hepler WT, Henderson RA;
 XX WPI; 2001-639232/73.

PT New human prostate-specific polypeptides and polynucleotides useful for
 the diagnosis and treatment of cancer, especially prostate cancer -
 XX Example 17; Page 541-542; 579pp; English.

CC The invention relates to isolated prostate-specific
 CC polynucleotides, polypeptides, fusion proteins of the polypeptides,
 CC antibodies raised against the polypeptides (or antigenic epitopes
 CC derived from them) and antigen-presenting cells expressing the
 CC polypeptides. The antibodies are useful for detecting the presence of
 CC cancer, especially prostate cancer. The polypeptides, polynucleotides and
 CC the antigen-presenting cells are useful for stimulating and/or expanding
 CC T cells specific for a tumour protein, and for inhibiting the development
 CC of cancer, especially prostate cancer. Compositions comprising the
 CC polynucleotide and/or polypeptide are useful for stimulating an immune
 CC response, and for treating cancer. The oligonucleotide is useful for

CC detecting cancer. The present sequence is M. tuberculosis Ra12
 CC antigen (or fragment) used to make a fusion protein comprising a
 CC prostate specific polypeptide of the invention.

SQ Sequence 132 AA;

Query Match 99.3%; Score 670; DB 22; Length 132;
 Best Local Similarity 99.2%; Pred. No. 7.4e-59;
 Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGGFAIPICGAMATAAGQIRSGGSPVTHIGTATLGLGVNDNNGAR 60
 |||||||
 Db 1 taasdnfqlsggggfaipigamataagqlrsggspvthlglglgvndnngar 60
 |||||||
 QY 61 VORVVGSAFPAASLIGSTGDIITAVDGPINSATAMADALNGHHPGDVISTWTKSGGTR 120
 |||||||
 Db 61 vgrvvgsapaasligstgdiitavdgaplnsatamadalnghpgdvisvnmwtksggtr 120
 |||||||
 QY 121 TGNVTLAEGPPA 132
 |||||||
 Db 121 tgnvtlaegppa 132

RESULT 15
 ID AAM01253 standard; protein; 132 AA.

AC AAM01253;
 04-OCT-2001 (first entry)

DE Mycobacterium tuberculosis antigen Ra12 amino acids.
 KM Human; prostate cancer; prostate-specific; diagnosis; vaccine;
 cytostatic; gene therapy; metastasis.

OS Mycobacterium tuberculosis.
 PN WO200151633-A2.
 PD 19-JUL-2001.

PF 16-JAN-2001; 2001MO-US01574.
 XX 14-JAN-2000; 2000US-0483672.
 PR (CORI-) CORIXA CORP.

XX Xu J, Dillon DC, Mitcham JL, Harlocker SL, Reed SG;
 PI Kalos MD, Fanger GR, Retter MW, Stolk JA, Skeiky YAW;
 PI Wang A, Meagher MJ;
 DR WPI; 2001-425873/45.

PT New polynucleotide encoding a prostate-specific protein, for
 diagnosing, monitoring and treating prostate cancer in a patient and
 PT for use in vaccines -
 XX Example 17; Page 492; 543pp; English.

CC The present invention describes polynucleotide sequences (I) which encode
 CC prostate-specific proteins (II). (I) and (II) have cytostatic activity,
 CC and can be used in vaccine production and gene therapy. (I), (II),
 CC antibodies to (II), fusion proteins comprising (II), and isolated
 CC T cells prepared using (I) or (II) are used to treat cancer in a patient.
 CC (I) and the antibodies are also used in the detection of cancer in a
 CC patient. The cancer that is diagnosed or treated is particularly
 CC prostate cancer. (I) and (II) can be used in vaccines. The antibodies or
 CC (I) can be used for monitoring the progression of cancer in a patient.
 CC (I) and (II) can also be used to improve diagnostic and therapeutic
 CC methods for prostate cancer. They can indicate the level of metastasis
 CC as well as the prostate volume. AAH93357 to AAH93944 and AAM01115 to

CC AAM01318 represent polynucleotide and amino acid sequences used in the
CC exemplification of the present invention.

XX
SQ Sequence 132 AA;

Query Match 99.3%; Score 670; DB 22; Length 132;

Best Local Similarity 99.2%; Pred. No. 7.4e-59;

Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TAAADNFQISOGGCGFAIPISGAMATAGQIRSGGSPYVHIGPTAFILGLGVNDNNGGAR 60

Db 1 taasdnfglsqgggfaipigamalaaggirsqgssptvhiplafilyvvdnnnggar 60

OY 61 VQRVVGSAFPAASLGISTGDPVITAVDGAPINSATAMADALNGHHGPDVISTWQTKSGCTR 120

Db 61 vqrvvgsapaaslgistgtgvtavdgapinsatamadalinghpgdvistwnqtksgctr 120

OY 121 TGNVTLAEGPPA 132

Db 121 tgnvclaeppa 132

Search completed: August 6, 2002, 08:15:50
Job time: 450 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 08:12:40 ; Search time 19.71 Seconds
(without alignments) 643.521 Million cell updates/sec

Title: US-09-684-215A-4

Perfect score: 675
Sequence: 1 TAAADNFQLSQGGGFAIRP.....QTKSGTRKGNVTLAEGPPA 132

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR_71.*

1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	DB ID	Description
1	675	100.0	355	2	F70983
2	474.5	70.3	361	2	S47170
3	459.5	68.1	354	2	A87242
4	199.5	29.6	464	2	C70821
5	188.5	27.9	452	2	H86930
6	188.5	27.9	452	2	T45448
7	139	20.6	394	2	S74643
8	139	20.6	407	2	AG2150
9	138.5	20.5	362	2	T35287
10	133.5	19.8	408	2	H86891
11	131	19.4	441	2	E75357
12	128.5	19.0	514	2	A82581
13	122.5	18.1	474	2	F83550
14	119.5	17.7	203	2	T35866
15	117.5	17.4	530	2	F87590
16	113	16.7	393	2	E95261
17	113	16.7	397	2	B98127
18	113	16.7	513	2	I40060
19	113	16.7	513	2	AD3418
20	113	16.7	523	2	A97479
21	113	16.7	523	2	A12696
22	112	16.6	429	2	AD1894
23	111	16.4	455	2	C91142
24	111	16.4	455	2	F85987
25	111	16.4	499	2	B81914
26	109	16.1	455	2	A80909
27	108	16.0	348	2	H96956
28	108	16.0	472	2	C87408
29	107.5	15.9	452	2	S77538

30	106.5	15.8	355	1	JC6052	trypsin-like prote
31	106.5	15.8	355	2	D91142	proteinase [import
32	106.5	15.8	355	2	G85987	proteinase [import
33	106.5	15.7	455	2	JC6051	trypsin-like prote
34	105	15.6	475	1	A15337	heat shock protein
35	104.5	15.5	362	2	A10433	proteinase (EC 3.4
36	102.5	15.2	474	2	S45229	proteinase DO (EC
37	102.5	15.2	474	2	E83500	proteinase DO (EC
38	102.5	15.2	474	2	E90649	proteinase DO (EC
39	102	15.1	475	2	AC0528	proteinase DO precu
40	100.5	14.9	453	1	B70426	periplasmic serine
41	100.5	14.9	459	2	F72359	periplasmic serine
42	100	14.8	398	2	B71284	probable periplasm
43	100	14.8	652	2	E97857	cell surface anty
44	99.5	14.7	224	2	F70048	serine proteinase
45	98.5	14.6	481	2	D82826	heat shock protein

ALIGNMENTS

RESULT 1
F70983
probable serine proteinase pepa - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_rev150 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: F70983
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
A: Reference number: A70500; MUID: 98295987
A: Accession: F70983
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1355 <COL>
A: Cross-references: GB:296071, GB:AL123456; NID:93242254; PIDN:CA809453.1; PID:921819
A: Experimental source: Strain H37RV
A: Genetics:
A: Gene: pepa
A: Superfamily: Escherichia coli trypsin-like proteinase

Query Match 100.0%; Score 675; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 9.2e-49;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAAADNFQLSQGGGFAIRPGQAMATAGQIRSGGSGPTVHIGTFAELGIVVDNNGGAR 60
Db 224 TAAADNFQLSQGGGFAIRPGQAMATAGQIRSGGSGPTVHIGTFAELGIVVDNNGGAR 283

QY 61 VQRVGAPAPASLGISTGDIYTAVDGAPINSATPMADALNGHHPGDIYSTWQTKSGGR 120
Db 284 VQRVGAPAPASLGISTGDIYTAVDGAPINSATPMADALNGHHPGDIYSTWQTKSGGR 343

QY 121 TGNVTLAEGPPA 132
Db 344 TGNVTLAEGPPA 355

RESULT 2
S47170
hypothetical protein 34K - Mycobacterium paratuberculosis
C:Species: Mycobacterium paratuberculosis
C:Date: 20-Feb-1995 #sequence_rev150 20-Feb-1995 #text_change 22-Oct-1999
C:Accession: S47170
R: Cameron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.
Submitted to the EMBL Data Library, June 1993
A: Description: Isolation and characterisation of a 34Kda protein of Mycobacterium par
A: Reference number: S47170
A: Accession: S47170

RESULT 6
T45448
probable serine proteinase (EC 3.4.21.-) MLCB373.28 [similarity] - Mycobacterium leprae

C:Species: Mycobacterium leprae
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 06-Oct-2000
C:Accession: T45448
R:James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, February 1998

A:Reference number: 222967
A:Accession: T45448
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-452 <JAN>
A:Cross-references: EMBL:AL035500; PIDN:CAB36690.1
A:Experimental source: cosmid L373
C:Genetics:
A:Note: MLCB373.28
C:Keywords: hydrolase; serine proteinase
F:182,224,305/Active site: His, Asp, Ser #status predicted

Query Match 27.9%; Score 188.5; DB 2; Length 452;
Best Local Similarity 39.8%; Pred. No. 2,66-08;
Matches 51; Conservative 19; Mismatches 51; Indels 7; Gaps 3;

OY 2 AASDNFQSLSGSGGFAIPICGAMAIAGQIRSGGSPYHIGPTAFGLGVVDNNGN-GAR 60
DB 328 ADSGASGSLGIPDQAKRIADELISTG--KATH---ASLGVAVATDKGTGPAK 381
OY 61 VORVGSAPASLIGISTGDTAVDGAIPINSATAMADALNGHHPGDVISTWOTKSGCTR 120
DB 382 VMDVVGAGAAAVAPKGVTLKVDRLISSADALVAVRKAPDKVSLTYQDQSSSR 441
OY 121 TGNVTIAE 128
DB 442 TQVVTILGK 449

RESULT 7
S74643
proteinase hbaA (EC 3.4.-.-) - *Synechocystis* sp. (strain PCC 6803)
N:Alternate names: protein g111679
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S74643
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* s.

A:Reference number: S74322; MUID:97061201
A:Accession: S74643
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-394 <KAN>
A:Cross-references: EMBL:D90900; GB:AB001339; NID:g1651768; PIDN:BAAL6795.1; PID:g165186
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: hbaA
C:Superfamily: proteinase hbaA
C:Keywords: hydrolase; proteinase

Query Match 20.6%; Score 139; DB 2; Length 394;
Best Local Similarity 34.6%; Pred. No. 0.00028;
Matches 45; Conservative 10; Mismatches 45; Indels 32; Gaps 5;

OY 13 GGGFAIPICGAMAIAGQIRSGGSPYHIGPTAFGLGV---VDNNGN-----G 58
DB 261 GIGFAIPIDQAKRIADGTVPHPIG-VQMMNITVDQAOQNNRNNSPFIPEVDG 319
OY 59 ARQVRVGSAPASLIGISTGDTAVDGAIPINSATAM-----ADALNGHHP 104
DB 320 ILVRVPLGTPARAGIRGDIVIAVDGTPISDGAIRLQIVEQAGLNKALKIDLGRDR 379

OY 105 GDVISTWOT 114
DB 380 ---LSLTVQF 386

RESULT 8
AG2150
serine proteinase [imported] - *Anabaena* sp. (strain PCC 7120)
C:Species: *Anabaena* sp.
A:Note: *Anabaena* sp. (strain PCC 7120) is a synonym of *Nostoc* sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AG2150
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuriitz, T.; Sasamoto, S.; Watanabe, A.; Irigu
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG2150
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW74457.1; PID:g17131851; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2758

Query Match 20.6%; Score 139; DB 2; Length 407;
Best Local Similarity 33.3%; Pred. No. 0.00029;
Matches 45; Conservative 17; Mismatches 47; Indels 26; Gaps 5;

OY 13 GGGFAIPICGAMAIAGQIRSGGSPYHIGPTAFGLGV-----VDNNG----- 56
DB 275 GIGFAIPIDQAKRIADGTVPHPIG---PYLGVQMTLPPELAQNNIDPNSAFI 328
OY 57 ---NGARQVRVGSAPASLIGISTGDTAVDGAIPINSATAMADALNGHHPGDVISTWQ 113
DB 329 PEVNGVIVIRVPSAPANNIRGDIVILQVDGQITATQQLQNVENSRLGQALQVRLQ 388
OY 114 TKSGGTRGNVTIAE 128
DB 389 -RGNQTOOLSVRTAE 402

RESULT 9
T35287
probable secreted proteinase - *Streptomyces coelicolor*
C:Species: *Streptomyces coelicolor*
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 21-Jan-2000
C:Accession: T35287
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: 221574
A:Accession: T35287
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-362 <SEE>
A:Cross-references: EMBL:AL096872; PIDN:CAB51255.1; GSPDB:GN00070; SCOEDB:SC5F7.30
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC5F7.30
C:Superfamily: *Escherichia coli* trypsin-like proteinase

Query Match 20.5%; Score 138.5; DB 2; Length 362;
Best Local Similarity 35.5%; Pred. No. 0.00028;
Matches 43; Conservative 14; Mismatches 55; Indels 9; Gaps 4;

OY 13 GGGFAIPICGAMAIAGQIRSGG---GSPYHIGPTAFGLGVVDNNGN---GARQVRVGS 67
DB 245 GIGFAIPASMTYVAGQIVRGKVTDSGRAALGTTA---RTVDDSDYRAGAAVVEVSDG 301
OY 68 APAASLIGISTGDTAVDGAIPINSATAMADALNGHHPGDVISTWOTKSGTRGNVTIA 127

Db 302 GAADAGLRGVDYVNIKGTDTTITSLSEALASMRPGDKTQVY-TRCKERTAEVTLG 360
 QY 128 E 128
 Db 361 E 361

RESULT 10

H86891

exported serine proteinase (EC 3.4.21.-) [imported] - Lactococcus lactis subsp. lactis
 C:Species: Lactococcus lactis subsp. lactis
 C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001

C:Accession: H86891
 R:Polotkin, A.; Winkler, P.; Mager, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Ehrlich

Genome Res. 11, 731-753, 2001
 A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis s

A:Reference number: A86625; MUID:21235186; PMID:11337471
 A:Accession: H86891

A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-408 <STO>
 A:Cross-references: GB:AE005176; PID:g12725194; PIDN:AAK06234.1; GSPDB:GN00146

A:Experimental source: strain 114403
 C:Genetics:

A:Gene: htra

C:Keywords: hydrolase; serine proteinase

Query Match 19.8%; Score 133.5; DB 2; Length 408;
 Best Local Similarity 31.1%; Pred. No. 0.00083;

Matches 41; Conservative 20; Mismatches 48; Indels 23; Gaps 5;

QY 13 GGGFAIPICGAMAIAGQIRSGGSPVHIGPFAIGLVVD-----NNGN----- 57
 Db 270 GGGFAIPICGAMAIAGQIRSGGSPVHIGPFAIGLVVD-----NNGN----- 57

QY 58 -GARVORVGSAPASLSTIGDVTAVDGPINSATAMADALNGHHDPDVISVTQTKS 116
 Db 324 GGVVSVSVSGSLPEASAGLAKADVDITKVDVSTDLQSLALYSHINDIVKVTYY-RD 382

QY 117 GGTTRGNVTLA 128
 Db 383 GKSNTADVTKSK 394

RESULT 11

E75357

probable periplasmic serine proteinase Do - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: E75357

R:White, O.; Eelsen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.

A:Reference number: A75250; MUID:20036896
 A:Accession: E75357

A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-441 <MHT>

A:Cross-references: GB:AE002017; GB:AE000513; NID:96459527; PIDN:AAF11312.1; PID:9645953

A:Experimental source: strain R1
 C:Genetics:

A:Gene: DR1756
 A:Map position: 1
 C:Superfamily: proteinase hhob

Query Match 19.4%; Score 131; DB 2; Length 441;
 Best Local Similarity 29.0%; Pred. No. 0.0015;
 Matches 47; Conservative 24; Mismatches 43; Indels 48; Gaps 6;

QY 6 NFQISGGQ-----GFAIPICGAMAIAGQIRSGGSPVHIGPFAIGLVVDN---- 55

Db 277 NQIILTGAGGAGAGGFAIPINTVKKRLPOLQAGKG-----GVSPSPSLGVFSDLSL 330

QY 56 -----GNGARVORVGSAPASLIG-----IST-GDVTAV 84

Db 331 PPQOLKAAGLPSGALLQKVPSPAAAGLRGNNKLSLPSAGCSISTDGLITAV 330

QY 85 DGAPINSATAMADALNGHHDPDVISVTQTKSGGTRFGNVL 126

Db 391 NGQPLEDAGSLQEAFLVATGEGQPLRTVR-RGKTRREVTL 431

RESULT 12

A82581

periplasmic proteinase XF2241 [imported] - Xylella fastidiosa (strain 9a5c)
 C:Species: Xylella fastidiosa

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
 C:Accession: A82581

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
 Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
 A:Reference number: A82515; MUID:20365717

A:Note: for a complete list of authors see reference number A59328 below
 A:Accession: A82581

A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-514 <STM>
 A:Cross-references: GB:AE004037; GB:AE003849; NID:g9107394; PIDN:AAF5040.1; GSPDB:GN

A:Experimental source: strain 9a5c
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Accencio, M.; Alvarenga, R.

Blanes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre
 as-Neto, E.; Docena, C.; El-Doroty, H.; Facincani, A.P.; Ferreira, A.D.S.

submitted to Genbank, June 2000
 A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Fr

J.D.; Jungueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
 chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri,

Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, F.R.; da Silva Jr., W.A.; da Silv

M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
 A:Reference number: A59328

A:Contents: annotation
 C:Genetics:

A:Gene: XF2241
 C:Superfamily: Helicobacter serine proteinase

Query Match 19.0%; Score 128.5; DB 2; Length 514;
 Best Local Similarity 38.7%; Pred. No. 0.0028;

Matches 43; Conservative 11; Mismatches 46; Indels 11; Gaps 4;

QY 4 SDNEQLSQG--GGGFAIPICGAMAIAGQIRSGG-----GSPVHIGPFAFL---GLGVVDN 54

Db 262 SQPSASGAGYMSIFALPIMLINAAGQIRKTKGVORSMVGVEIGPLDLAKAGGLTSPS 321

QY 55 NGNGARVORVGSAPASLSTIGDVTAVDGPINSATAMADALNGHHPG 105

Db 322 --KGALVNNIPHPSPAKAGIEVDVIRSVNGKVISFSDDLPLIGMPPG 370

RESULT 13

F83550

serine proteinase MucD precursor PA0766 [imported] - Pseudomonas aeruginosa (strain P
 C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: F83550

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidig, K.; L

.; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000

A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: F8350
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-474 <STO>
A:Cross-references: GB:AE004511; GB:AE004091; NID:9946646; PIDN:AA604155.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: mucD; PA0766

Query Match 18.1%; Score 122.5; DB 2; Length 474;
Best Local Similarity 30.4%; Pred. No. 0.008;
Matches 42; Conservative 22; Mismatches 51; Indels 23; Gaps 5;
QY 4 SDNFOLSG--GQFAIPDIGAMAIAGQIRSGSPVYHIGPT--AF-----IGLVVDNNGNGARVQR 63
DB 233 SQIFRSGFGLSFAPIDVAVLNADLKKAG--KVSRG--WLGVIQEVNKKDLAES 286
QY 56 -----GNGARVQRVYGSAPASLIGSTGDIYAVDGAIPNSATAMADALNGHHPGDVYSV 110
DB 287 FGLDKPSGALVAQLVEDGCPAKGGLQGVGVLTSLNGSINSADLPHLVGMKRGDKINL 346
QY 111 ----TWQTKSGGTRTGNV 124
DB 347 DVIRNGQRKSLSMVAGSL 364

RESULT 14 735866
probable integral membrane protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence.revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: J35866
R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, April 1999
A:Reference number: 221591
A:Accession: J35866
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-203 <SAU>
A:Cross-references: EMBL:AL049727; PIDN:CAB41567.1; GSPDB:GN00070; SCQDB:SC9B1.21
C:Genetics:
A:Experimental source: strain A3(2)
A:Gene: SCQDB:SC9B1.21

Query Match 17.7%; Score 119.5; DB 2; Length 203;
Best Local Similarity 32.1%; Pred. No. 0.0057;
Matches 36; Conservative 15; Mismatches 56; Indels 5; Gaps 1;
QY 12 GGCGFAIPDIGAMAIAGQIRSGSPVYHIGPT-----AFGLGVVDNNGNGARVQRVY 66
DB 80 GGAGGQATGTATGTGRAPHRPSAGRSFSPAPACATLGEAVDDDEKPARVGVHV 139
QY 67 SAPASLIGSTGDIYAVDGAIPNSATAMADALNGHHPGDVYSVTWQTKSGG 118
DB 140 PGPGYAAGLVAGVDLAVGTTRVDSATDLAAHVARAGPGKEVKLTIVRHRSRG 191

RESULT 15 F87590
serine proteinase Htra [imported] - *Caulobacter crescentus*
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence.revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: F87590
R:Nierman, W.C.; Feldblum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolod
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of *Caulobacter crescentus*.
A:Reference number: A87249; MUID:21173698; PMID:11259647

A:Accession: F87590
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-530 <STO>
A:Cross-references: GB:AE005673; NID:g13424352; PIDN:AAK24722.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2758
C:Superfamily: Helicobacter serine proteinase

Query Match 17.4%; Score 117.5; DB 2; Length 530;
Best Local Similarity 31.6%; Pred. No. 0.023;
Matches 42; Conservative 18; Mismatches 50; Indels 23; Gaps 5;
QY 13 GQFAIPDIGAMAIAGQIRSGSPVYHIGPT--AF-----IGLVVDNNGNGARVQR 63
DB 283 GIGFALPAEYVAGVAKQLENGKVVVRGIVGVSIMARNAEALGMSDV---KGAIVAS 338
QY 64 VGSAPASLIGSTGDIYAVDGAIPNSATAMADALNGHHPGDVYSVT--WQT 114
DB 339 VVPGGPAKAKGLLPDILVAVNGVKISDSSELTREVSKARPGETIKVSIIRDGKPRIVDV 398
QY 115 KSGGTRTGNVTLA 127
DB 399 KS-GTPRESSLA 410

Search completed: August 6, 2002, 08:16:53
Job time: 253 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 08:15:10 ; Search time 10.04 Seconds
(without alignments)
509.062 Million cell updates/sec

Title: US-09-684-215a-4

Perfect score: 675
Sequence: 1 TAAADNFQSLSGGQGFALPI.....QTKSGGTGTGNTVLAEGPFA 132

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	133.5	19.8	408	HTRA_LACLA	Q91A06 Lactococcus
2	125	18.5	413	HTRA_LACHE	Q9Z4H7 Lactobacilli
3	118	17.5	504	DEGP_RHIME	Q52894 Lactobacium m
4	113	16.7	513	DEGP_BRUAB	Q44597 Brucella ab
5	106.5	15.8	355	DEGS_ECOLI	P31137 Escherichia
6	106	15.7	455	DEGO_ECOLI	P39089 Escherichia
7	105	15.6	475	DEGP_SALTY	P26982 Salmonella
8	102.5	15.2	474	DEGP_ECOLI	P09376 Escherichia
9	101.5	15.0	503	DEGP_BARE	P54925 Bartonella
10	97	14.4	497	DEGP_CHLTR	P18584 Chlamydia t
11	94.5	14.0	497	DEGP_CHLMU	Q9P197 Chlamydia m
12	94.5	14.0	957	Y278_MYCTU	P56877 Mycobacteri
13	92.5	13.7	466	HTOA_HAEIN	P45129 Haemophilus
14	92	13.6	448	DEGP_ARATH	Q91U10 Arabidopsis
15	92	13.6	478	DEGP_BUCAI	P57332 Buchnera ap
16	89.5	13.3	905	Z03_MOUSE	Q9QXY1 Mus musculu
17	87	12.9	864	ELIS_RAT	Q99372 Rattus norv
18	86	12.7	762	P15_CHICK	Q98917 Gallus gall
19	86	12.7	2249	OMPA_RICRI	P15931 Rickettsia
20	85	12.6	515	Y140_MYCTU	Q05354 Mycobacteri
21	85	12.6	2021	OMPA_RICCN	Q52657 Rickettsia
22	83	12.3	488	DEGP_CHLPP	Q92660 Chlamydia p
23	83	12.3	510	MURF_MYCTU	Q06220 Mycobacteri
24	82	12.1	389	HMUX_DROME	P02834 Drosophila
25	81.5	12.1	219	VG36_BPT2	P07066 Bacterioph
26	81.5	12.1	619	ELF1_HUMAN	P22519 Homo sapien
27	81.5	12.1	714	CDG1_PALMA	P04830 Penicillium
28	81	12.0	354	ODPA_ZYMO	O66112 Zymomonas m
29	81	12.0	430	TPSN_CHICK	O73895 Gallus gall
30	80	11.9	382	YN28_MYCTU	P18184 Mycobacteri
31	79.5	11.8	292	PPNK_ECOLI	P58057 Escherichia
32	79.5	11.8	292	PPNK_ECOLI	P37766 Escherichia
33	79.5	11.8	1286	AIDA_ECOLI	Q03155 Escherichia

34	79	11.7	253	1	HMUX_DROME	P05048 drosophila
35	79	11.7	257	1	ECHE_MYCLE	O07137 mycobacteri
36	78.5	11.6	351	1	HYPE_ALCEU	P31905 alcaligenes
37	78	11.6	159	1	19KD_MYCTU	P15670 mycobacteri
38	78	11.6	801	1	Y747_MYCTU	O53810 mycobacteri
39	77.5	11.5	478	1	DEGP_BUCAP	O85291 buchnera ap
40	77.5	11.5	479	1	XVIR_LACPE	P96792 lactobacilli
41	77.5	11.5	933	1	Z03_HUMAN	O95049 homo sapien
42	77	11.4	396	1	YD61_MYCTU	O11031 mycobacteri
43	77	11.4	543	1	YD61_DROME	P16375 drosophila
44	77	11.4	746	1	YD61_DROME	P16376 drosophila
45	77	11.4	778	1	YQ34_MYCTU	P71933 mycobacteri

ALIGNMENTS

RESULT 1
HTRA_LACLA STANDARD; PRT; 408 AA.
AC Q9LA06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Serine protease do-1like htra (EC 3.4.21.-) (HtraLl).
GN HTRA OR IL2136
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=IL1403;
RX MEDLINE=20177820; PubMed=10712686;
RA Poquet I., Saint V., Seznec E., Simoes N., Bolotin A., Gruss A.;
RT "Htra is the unique surface housekeeping protease in Lactococcus
RT lactis and is required for natural protein processing.";
RL Mol. Microbiol. 35:1042-1051(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Winkler P., Mauger S., Jallion O., Malarme K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
CC -1- FUNCTION: Degrades abnormal exported proteins. Needed for the pro-
CC peptide processing of a natural pro-protein and for maturation of
CC A native protein. Responsible for the housekeeping of exported
CC proteins.
CC -1- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGO/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AF155705; AAF61294.1; -
CC EMBL: AE006442; AAK06234.1; -
CC InterPro: IPR001478; PDZ
CC InterPro: IPR001940; ProteaseZC.
CC InterPro: IPR001254; Trypsin.
CC Pfam: PF00595; PDZ; 1.
CC Pfam: PF00089; Trypsin; 1.
CC PRINTS: PR00834; PROTEASES2C.
CC SMART: SMO0228; PDZ; 1.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U31512; AAC43669.1; ALT_INT.
DR EMBL: AL591785; CAC45593.1; -.
DR MEROPS: S01.273; .
DR InterPro: IPR001476; PDZ.
DR InterPro: IPR001940; Protease2C.
DR -----

```

DR InterPro: IPRO00126; Ser-proteas_V8.
DR pfam: PF000595; PDZ: 2.
DR Pfam: PR00089; trypsin_1.
DR PRINTS: PR00834; PROTEASES2C.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SMO0228; PDZ: 2.
KW POSITIVE: P550106; PDZ: 2.
KM Hydrolyase; Serine protease; Signal; Periplasmic; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 504
FT DOMAIN 113 286
FT DOMAIN 287 378
FT DOMAIN 378 491
FT ACT_SITE 401 491
FT ACT_SITE 140 140
FT ACT_SITE 170 170
FT ACT_SITE 244 244
FT ACT_SITE 244 244
FT CONFLICT 14 15
FT CONFLICT 39 147
FT CONFLICT 147 147
PEPANYVDVSPAVSVRVQAREVRVSDSSNFTDPRGGRGF
EDPEDHPLRRFPREFAPENDRADRWDRRGPRGGRLRP
RAOGEFFETDEGYLVTNHHVDSGA -> AVSPWSTPF
RRRSSPSACRHVNASTMTKATSPISIAAGSRICRTIKCG
VSSANSLRYKMTVPVIGATPAVAVKVAKVSVRGKKAPASSP
KYTVSSPTTSSPTART (IN REF. 1).
KSADYDIKVITNNMKGKRSALFQIEAOGSSRVALPIIYG
-> NROTTPSR (IN REF. 1).
SQ SEQUENCE 504 AA; 53035 MW; D7EB2BB9981EA23C CRC64;

Query Match 17.5%; Score 118; DB 1; Length 504;
Best Local Similarity 30.9%; Pred. No. 0.016;
Matches 42; Conservative 14; Mismatches 62; Indels 18; Gaps 6;

QY 7 FOLSGGGG--FAIPICOMAIAGQIRSG----GSPTVHIIGPPA-----FLGLGVYDNN 55
      | | | | | | | | | | | : | | | | | : | | |
DB 263 FSPSGGNVGIFAIPAASAKDYVDLSLTKDGVSRCGWGVQIQPTKDIABSLGI-----SE 318
      | | | | | | | | | | | : | | | | | : | | |
QY 56 NGCARVCRVVASAPAAASLGISTGDVTAVAGCAPINSATAMDALNGHPGPVISVT-WQT 114
      | | | | | | | | | | | : | | | | | : | | |
DB 319 ANGALVVEPQASGEKEKAGIKNGDVITALNGEVPKDPDLARVALRPGSTAELVTIM-- 376
      | | | | | | | | | | | : | | | | | : | | |
QY 115 KSGGRTGNTVLAEGP 130
      | | | | | | | | | | | : | | | | | : | | |
DB 377 RSGKSETVNLIGTLF 392
      | | | | | | | | | | | : | | | | | : | | |

RESULT 4
ID DEGP_BRUAB STANDARD; PRT; 513 AA.
AC 044597;
DC 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-1like precursor (EC 3.4.21.-).
GN DEGP OR HTRA.
OS Brucella abortus.
OC Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;
CC Brucellaceae: Brucella.
OX NCBI_TaxId=235;
      [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2308;
RX MEDLINE=95165990; PubMed=7861951;
RA Tatum F.M., Cheville N.F., Morfitt D.;
RT "Cloning, characterization and construction of htra and htra-like
RL mutants of Brucella abortus and their survival in BALB/c mice.";
RL Microb. Pathog. 17:23-36(1994).
CC -1- SUBCELLULAR LOCATION: periplasmic (potential).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY SZC; ALSO KNOWN AS THE
CC DEGP/DEGO/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```

CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL: U07352; AAA70164.1; -.
DR      InterPro: IPR001478; PDZ.
DR      InterPro: IPR001940; Protease2C.
DR      InterPro: IPR001254; Trypsin.
DR      Pfam: PF00595; PDZ_2.
DR      Pfam: PF00089; Trypsin; 1.
DR      PRINTS: PRO0834; PROTEASESZC.
DR      SMART: SM00228; PDZ_2.
DR      PROSITE: PS0106; PDZ_2.
KW      Hydrolyase, Serine protease; signal; periplasmic.
FT      SIGNAL 1 25
FT      CHAIN 26 513
FT      DOMAIN 125 299
FT      DOMAIN 300 391
FT      DOMAIN 414 500
FT      ACT_SITE 152 152
FT      ACT_SITE 182 182
FT      ACT_SITE 257 257
FT      ACT_SITE 513 AA; 53483 MW; DELCEP1959472806 CXC64;
SQ      SEQUENCE 513 AA; 53483 MW; DELCEP1959472806 CXC64;

Query Match 16.7%; Score 113; DB 1; Length 513;
Best Local Similarity 33.3%; Pred. No. 0.041;
Matches 45; Conservative 11; Mismatches 63; Indels 16; Gaps 5.

OY      7 FOLSGGGG--FAIRPGQMAINAGQIRSGG-----GSPYHNIGP-----TAFICIGVNDNN 55
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      276 FSPSGGSGVGIAIPSPSTAKOVVDOLIKKSGVERGNIQVQIDQVTKDIASIGLA-----E 331
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      56 GNGARQVRVVGSPAPASLIGTSDVITAVDGAAPINATAMADALNCHNPDIISVTWQTK 115
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      332 EKGAIYASQDDGSPAKAKIGKAGDVTITVNGEIVQDPRLARVAVIARGEKALYVWRK 391
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY      116 SGGTRGTNTLAEGP 130
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB      392 NKAEI-NVTIAMP 405
          | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
DEGS.ECOLI
ID      DEGS_ECOLI STANDARD; PRT; 355 AA.
AC      P31137;
DT      01-JUL-1993 (Rel. 26, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Protease deags precursor (EC 3.4.21.-).
GN      DEGS OR HCOB OR HTRP OR B3235 OR Z4594 OR ECS4108.
OS      Escherichia coli, and
OS      Escherichia coli O157:H7.
OC      Escherichia coli O157:H7.
OC      Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC      Escherichia
OX      NCBI_TaxID=562, 83334.
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / W3110;
RA      Bass S., Gu O., Goddard A.;
RL      Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=K12 / W3110;
RX      MEDLINE=96165272; PubMed=8576051;
RT      Waller P.R., Sauer R.T.;
RT      "Characterisation of dego and degs, Escherichia coli genes encoding
RT      homologs of the DegP protease."
RL      J. Bacteriol. 178:1146-1153(1996).
RN      [3]

```

RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RA MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=1120551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dialanta E.T., Potamodis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.";
 RL Nature 409:529-533(2001).
 [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
 O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 [6]
 RP PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
 RX MEDLINE=88105815; PubMed=3322223;
 RA Vogel R.F., Entlian K.-D., Mecke D.;
 RT "Cloning and sequence of the mdh structural gene of *Escherichia coli*
 coding for malate dehydrogenase.";
 RL Arch. Microbiol. 149:36-42(1987).
 [7]
 RP IDENTIFICATION.
 RA Bazan J.F., Fleeterick R.J.;
 RT "Structural and catalytic models of trypsin-like viral proteases.";
 RL Semin. Virol. 1:311-322(1990).
 CC -1- SUBCELLULAR LOCATION: Periplasmic (potential).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 DEGP/DEGP/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U15661; AAC43993.1; -;
 DR EMBL: U32495; AAC44006.1; -;
 DR EMBL: U18997; AAA58037.1; -;
 DR EMBL: AE000402; AAC76267.1; -;
 DR EMBL: AE005551; AAG58363.1; -;
 DR EMBL: AP002564; BAB37531.1; -;
 DR EMBL: M24777; -: NOT_ANNOTATED_CDS.
 DR MEGS: S01.275; -;
 DR EcoGene: EG11652; -;
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF000595; PDZ.1.
 DR Pfam: PF00089; PDZ.1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ.1.

DR PROSITE: PS0106; PDZ.1.
 KW Hydrolase; serine protease; periplasmic; signal; complete proteome.
 FT SIGNAL 1 28
 FT CHAIN 29 355
 FT DOMAIN 281 326
 FT ACT_SITE 96 96
 FT ACT_SITE 126 126
 FT ACT_SITE 201 201
 FT CONFLICT 253 253
 FT CONFLICT 307 307
 SQ SEQUENCE 355 AA; 37581 MW; D091B4D65E8FE1CC CRC64;
 Query Match 15.8%; Score 106.5; DB 1; Length 355;
 Best Local Similarity 31.2%; Pred. No. 0.092;
 Matches 44; Conservative 10; Mismatches 64; Indels 23; Gaps 5;
 QY 6 NFOLSGGQ-----GFAIP-----IGGAMVAGQIRSGGSPYHIGPTFLGL 49
 Db 219 SFVKSNDGEIPREGIPFQATKIMDKLIRGKRYIRGIGIGR---EIAPLHAQGG 274
 QY 50 GVVNDNGGARVQVYVGSAPASLIGISTGDTITAVDGPINSATAMADALNGHPDVIS 109
 Db 275 GI--DLOGIIVNEVSPDGPAAVAGIQVNDLISVKNKPAISLETMDQVAELRPSVIP 332
 QY 110 VYMQTSGSGTRGCVTLASGP 130
 Db 333 VV-VMRDKQLTLQVYIQEYP 352
 RESULT 6
 DECO_ECOLI
 ID DECO_ECOLI STANDARD; PRT; 455 AA.
 AC P39099;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protease dego precursor (EC 3.4.21.-).
 GN DEGO OR HHOA OR B3234.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_Taxid=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / W3110;
 RA Bass S., Gu O., Goddard A.;
 RL Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.
 RC STRAIN-K12 / W3110;
 RX MEDLINE=96165272; PubMed=8576051;
 RA Waller P.R., Sauer R.T.;
 RT "Characterization of dego and degs, *Escherichia coli* genes encoding
 homologs of the DegP protease.";
 RL J. Bacteriol. 178:1146-1153(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 CC -1- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DEGP.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 DEGP/DEGO/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: U15661; AAC3992.1; -
 DR EMBL: U32495; AAC4005.1; -
 DR EMBL: U18997; AAA58036.1; -
 DR EMBL: AE000402; AAC76266.1; -
 DR MEROPS: S01.274; -
 DR SWISS-2DPAGE: P39099; COLI.
 DR Ecocore: EG12612; dego.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 DR Hydrolyse: Serine protease; periplasmic; signal; complete proteome.
 KW SIGNAL
 FT CHAIN 1 27
 FT DOMAIN 28 455
 FT DOMAIN 258 349
 FT ACT_SITE 109 109 PDZ 1.
 FT ACT_SITE 139 139 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 455 AA; 47205 MW; 6A050F93AC021C83 CRC64;

Query Match 15.7%; Score 106; DB 1; Length 455;
 Best Local Similarity 26.5%; Pred. No. 0.13; Indels 40; Gaps 5;
 Matches 36; Conservative 19; Mismatches 41;

QY 13 GGGFAIPGQAMIA-----GGIRSGGSPVNHGPTAFLGLGVNDNGN----- 57
 DB 241 GIGFAIPSNMAYFLAQOLIDFGEIKRG-----DLGIGKTEMSADIAKAFMLDV 288
 QY 58 --GARQVRVGSAPASLSIGSTGDVITAVDGAIPNSATAMADLNGHHPDVISVTWQK 115
 DB 289 QKRAFYSEVLPGGSGAKAGKADITSLNGKPLNSFALRLSRITATPEPT-----KVK 342
 QY 116 SGGTRTG-----NVTL 126
 DB 343 LGLLRNCKPLEVEYTL 358

RESULT 7
 DEGP_SALTY STANDARD; PRT; 475 AA.
 ID DEGP_SALTY
 AC P26982;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Protease do precursor (EC 3.4.21.-)
 GN DEGP OR HTRA OR PTD OR STM0209.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C5;
 RX MEDLINE=91251770; PubMed=1645840;
 RA Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,
 RA All T., Miller I., Hormaeche C.;
 RT "The role of a stress-response protein in Salmonella typhimurium
 RT virulence."
 RL Mol. Microbiol. 5:401-407(1991).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2."
 RL Nature 413:852-856(2001).
 CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
 CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED
 CC SPECIFICITY WITH HROA/DEGO.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- INDUCTION: BY HEAT SHOCK.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
 CC DEGP/DEGO/DEGS FAMILY.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: X54548; CAA38420.1; -
 DR EMBL: AE008704; AAL19173.1; -
 DR PIR: S15337; S15337.
 DR PIR: S21327; S21327.
 DR MEROPS: S01.273; -
 DR StyGene: SG10173; degp.
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ; 2.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PS50106; PDZ; 2.
 DR Hydrolyse: Serine protease; Heat shock; periplasmic; signal;
 KW Complete proteome.
 FT SIGNAL 1 26
 FT CHAIN 27 475
 FT DOMAIN 281 372
 FT DOMAIN 378 467
 FT ACT_SITE 132 132 PDZ 2.
 FT ACT_SITE 162 162 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 237 237 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 475 AA; 49315 MW; 86E65BF3C1A289F CRC64;

Query Match 15.6%; Score 105; DB 1; Length 475;
 Best Local Similarity 28.1%; Pred. No. 0.16; Indels 34; Gaps 4;
 Matches 38; Conservative 19; Mismatches 44;

QY 13 GGGFAIP-----IGQAMINGQIRSGGSPVNHGPTAFLGLGVNDNGN----- 57
 DB 264 GIGFAIPSNMAYKNTLSQWVEYGOVVRG-----ELGIMTELSELAKAMKV 309
 QY 58 ----GARQVRVGSAPASLSIGSTGDVITAVDGAIPNSATAMADLNGHHPDVISVTWQ 113
 DB 310 DAQRGAFVSQVMPNSAKAGIKAGDVITSLNGKPISSFALRAQVGMIPGSKISL-GL 368
 QY 114 TKSGGTRTGNVTLAE 128
 DB 369 LREGKATIVNLELQ 383

RESULT 8
 DEGP_ECOLI

```

ID DEGP_ECOLI STANDARD: PRT: 474 AA.
AC P09376; P15724;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR PPD OR B0161 OR Z0173 OR ECS0165.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89057448; PubMed=3057437;
RA Lipinska B., Sharma S., Georgopoulos C.;
RT "Sequence analysis and regulation of the htra gene of Escherichia
RT coli: a sigma 32-independent mechanism of heat-inducible
RT transcription.";
RL Nucleic Acids Res. 16:10053-10067(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Moti H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 2.4-4.1 min (110,917-193,643 bp) region.";
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=9742617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Laskerl D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RL [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobleck E.U., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsuno E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Tida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [7]
RP SEQUENCE OF 1-50 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=90323597; PubMed=2165018;

```

```

RA Quirk S., Bhattachar S.K., Bessman M.J.;
RT "Primary structure of the deoxyguanosine triphosphate
RT triphosphohydrolyase-encoding gene (dgt) of Escherichia coli.";
RL Gene 89:13-18(1990).
RN [8]
RP SEQUENCE OF 1-16 FROM N.A.
RX MEDLINE=90207273; PubMed=2157212;
RA Wurgler S.M., Richardson C.C.;
RT "Structure and regulation of the gene for dGTP triphosphohydrolyase
RT from Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744(1990).
RN [9]
RP CHARACTERIZATION AND SEQUENCE OF 27-39.
RX MEDLINE=90202693; PubMed=2180903;
RA Lipinska B., Zyllicz M., Georgopoulos C.;
RT "The Htra (DegP) protein, essential for Escherichia coli survival at
RT high temperatures, is an endopeptidase.";
RL J. Bacteriol. 172:1791-1797(1990).
RN [10]
RP IDENTITY OF HTRA AND PROTEASE DO.
RX MEDLINE=9122240; PubMed=2025286;
RA Seol J.H., Woo S.K., Jung E.M., Yoo S.J., Lee C.S., Kim K.J.,
RA Tanaka K., Ichihara A., Ha D.B., Chung C.H.;
RT "Protease Do is essential for survival of Escherichia coli at high
RT temperatures: its identity with the htra gene product.";
RL Biochem. Biophys. Res. Commun. 176:730-736(1991).
CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADS
CC ICIA, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGP.
CC -1- SUBUNIT: MULTIMERIC.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- MISCELLANEOUS: HTRA IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT
CC TEMPERATURES ABOVE 42 DEGREES CELSIUS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGO/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M36536; AAA23994.1; -
DR EMBL: X12457; CAA30997.1; -
DR EMBL: D26562; CAB20280.1; -
DR EMBL: AE000125; AAC73272.1; -
DR EMBL: U70214; AAB08591.1; -
DR EMBL: AE005192; AAG54465.1; -
DR EMBL: AP002550; BAB33588.1; -
DR EMBL: M29955; AAA23717.1; -
DR EMBL: M31772; AAA23680.1; -
DR PIR: S01899; S01899.
DR PIR: B35993; B35993.
DR MEROPS: S01.273; -.
DR SWISS-2DPAGE: P09376; COLI.
DR EcoGene: EG10463; degP.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR Pfam: PF000595; PDZ_2.
DR Pfam: PF00089; trypsin_1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ_2.
DR PROSITE: PS50106; PDZ_2.
KW Hydrolyase; Serine protease; Heat shock; Periplasmic; Signal;
KW Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 474 PROTEASE DO.
FT DOMAIN 280 371 PDZ 1.

```

```

FT DOMAIN 377 466 PDZ 2. RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 131 131 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 161 161 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 236 236 CHARGE RELAY SYSTEM (POTENTIAL).
FT CONFLICT 10 10 A -> R (IN REF. 1, 7 AND 8).
FT CONFLICT 46 46 E -> Q (IN REF. 7).
FT CONFLICT 192 192 A -> G (IN REF. 1).
FT CONFLICT 467 474 STYLLMO -> RHLPVNAVISLNPFLKTRGSPFYL (IN
FT SEQUENCE 474 AA: 49354 MW: 5482E596E7ABD65F CRC64;
SQ QUERY MATCH 15.2%; Score 102.5; DB 1; Length 474;
Best Local Similarity 30.4%; Pred. No. 0.25; 24; Indels 33; Gaps 3;
Matches 31; Conservative 14; Mismatches 24;

QY 13 GCGFAIP-----IGQAMAIAGQIRSGGSPYHIGPTAFLGLGVNDNGN----- 57
Db 263 GIGAFATSNMVKNTLSQMVEYGVRKG-----ELGIMGTSLNSLAKMKV 308

OY 58 -----GARQRYVGSAAPAASLGISTGDVTITAVDGAPINSATPM 95
Db 309 DAORGAFVSQVLPPNSSAKAGIKAGDVITSLNKRPISSFAAL 350

RESULT 9
DEGP_BARNE STANDARD; PRT: 503 AA.
AC P54925;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable periplasmic serine protease DO-like precursor (EC 3.4.21.-)
DE (Antigen htra).
GN DEGP OR HTRA.
OS Bartonella henselae (Rochalimaea henselae).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=38323;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HOUSTON-1.
RX MEDLINE=94299828; PubMed=8027347;
RA Anderson B., Sims K., Regniery R., Robinson L., Schmidt M.J.,
RA Gorai S., Hager C., Edwards K.;
RT "Detection of Rochalimaea henselae DNA in specimens from cat scratch
RT disease patients by PCR.";
RU J. Clin. Microbiol 32:942-948(1994).
CC -!- SUBCELLULAR LOCATION: Periplasmic (potential).
CC -!- SIMILARITY: BELONGS TO PERPLASMIC FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DEGO/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isdb.ch/announce/
CC or send an email to license@isdb-sib.ch).
CC CC
DR EMBL: L20127; AAA97430.1; -.
DR MEROPS: S01.273; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2c.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ_2.
DR Pfam: PF00089; trypsin.1.
DR PRINTS: PM00834; PROPEASES2C.
DR SMART: SM00228; PDZ_2.
DR PROSITE: PS50106; PDZ_2.
KW Hydrolase, Serine protease, Periplasmic, Signal,
FT SIGNAL 1 18 POTENTIAL.
```

[illegible]

```

CC -----
CC EMBL: AE001355; AAC68420.1; -
CC DR FMBL: M31119; AAA23116.1; -
CC DR InterPro: IPR001478; PDZ.
CC DR InterPro: IPR001940; Protease2C.
CC DR InterPro: IPR001254; Ser_proteas_V8.
CC DR Pfam: PF00595; PDZ. 2.
CC DR Pfam: PF00089; trypsin. 1.
CC DR PRINTS: PR00834; PROTEASES2C.
CC DR PRINTS: PR00839; V8PROTEASE.
CC DR SMART: SM00228; PDZ. 2.
CC DR PROSITE: PS0106; PDZ. 2.
CC KW Hydrolyase; Serine protease; Signal; Antigen; Complete proteome.
CC FT SIGNAL 1 16 POTENTIAL.
CC FT CHAIN 17 497 PROBABLE SERINE PROTEASE DO-LIKE.
CC FT DOMAIN 128 289 CATALYTIC.
CC FT DOMAIN 290 381 PDZ 1.
CC FT DOMAIN 394 485 PDZ 2.
CC FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
CC SQ SEQUENCE 497 AA; 53244 MW; 86A5E31BB84A38BA CRC64;

```

```

Query Match 14.4%; Score 97; DB 1; Length 497;
Best Local Similarity 29.9%; Pred. No. 0.73;
Matches 38; Conservative 18; Mismatches 57; Indels 14; Gaps 3;

```

```

QY 13 GCGFAITGQAMAIAGQIRSGGSPYVHIGPTAFGLGVND-----NGNGARVOR 63
DB 274 GIGFAIPISLMARKRVIDQLISDGVTRGFLGVT----LQPIDSELAACYKLEKYGALVTD 329
QY 64 VGSAPASASLGISTGDIVTAVDGAIPNSATAMADALNGHPGDIVISVTWQTKSGGRTGN 123
DB 330 VVGSPAPKAGKLRQEDIVAYVNGKEVESLSALRNATISLMMPGRV-VLKIRGKRTIEIP 388
QY 124 VTLAEGP 130
DB 389 VVTVOIP 395

```

```

RESULT 11
DEGP_CHLMTU STANDARD; PRT; 497 AA.
AC Q9PL97;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR TC0210.
OS Chlamydia muridarum.
OC Bacteria; Chlamydiales; Chlamydiaeae; Chlamydia.
OX NCBI_TaxID=83550;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / N199;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uettersack T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RA "Genome sequences of Chlamydia trachomatis MOPN and Chlamydia
RA pneumoniae AR39."
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
CC DEGP/DECO/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```

CC -----
CC EMBL: AE002288; AAF39082.1; -
CC DR TIGR: TC0210; -
CC DR InterPro: IPR001478; PDZ.
CC DR InterPro: IPR001940; Protease2C.
CC DR InterPro: IPR001254; Ser_proteas_V8.
CC DR Pfam: PF00595; PDZ. 2.
CC DR Pfam: PF00089; trypsin. 1.
CC DR PRINTS: PR00834; PROTEASES2C.
CC DR PRINTS: SM00228; PDZ. 2.
CC DR PROSITE: PS0106; PDZ. 2.
CC KW Hydrolyase; Serine protease; Signal; Complete proteome.
CC FT SIGNAL 1 16 POTENTIAL.
CC FT CHAIN 17 497 PROBABLE SERINE PROTEASE DO-LIKE.
CC FT DOMAIN 128 289 CATALYTIC.
CC FT DOMAIN 290 381 PDZ 1.
CC FT DOMAIN 394 485 PDZ 2.
CC FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
CC SQ SEQUENCE 497 AA; 53294 MW; B765F350AC66BBF CRC64;

```

```

Query Match 14.0%; Score 94.5; DB 1; Length 497;
Best Local Similarity 30.4%; Pred. No. 1.2;
Matches 31; Conservative 15; Mismatches 43; Indels 13; Gaps 2;

```

```

QY 13 GCGFAIPISLMARKRVIDQLISDGVTRGFLGVT----LQPIDSELAACYKLEKYGALVTD 329
DB 274 GIGFAIPISLMARKRVIDQLISDGVTRGFLGVT----LQPIDSELAACYKLEKYGALVTD 329
QY 64 VGSAPASASLGISTGDIVTAVDGAIPNSATAMADALNGHPG 105
DB 330 VVGSPAPKAGKLRQEDIVAYVNGKEVESLSALRNATISLMMPG 371

```

```

RESULT 12
Y278_MYCTU STANDARD; PRT; 957 AA.
ID Y278_MYCTU
AC P56877;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical PE-PGRS family protein RV0278c precursor.
GN RV0278C OR MT0291 OR MTW035.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltham T., Gentles S., Hamlin N., Holroyd S.,
RA Horsley T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.E., Taylor K., Whitehead S., Barrell B.G.,
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;

```

```

RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.,
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: AL021930; CAA17353.1; -.
CC EMBL: AE006936; AAK44511.1; ALT_INIT.
CC TIGR: MT0291; -.
CC TubercuList: Rv0278c; -.
CC InterPro: IPR000084; PE.
CC Pfam: PF00934; PE; 1.
CC DR Hypothetical protein; Repeat; Signal; Complete proteome.
CC KM SIGNAL 1 30 POTENTIAL.
CC FT CHAIN 1 957 HYPOTHEICAL PE-PGRS FAMILY PROTEIN
CC Rv0278c.
CC FT CONFLICT 40 40 M->I (IN REF. 2).
CC FT CONFLICT 158 163 MISSING (IN REF. 2).
CC FT CONFLICT 807 807 R->G (IN REF. 2).
CC SQ SEQUENCE 957 AA; 81905 MW; 71EBAD417FBA47C CRC64;

Query Match 14.0%; Score 94.5; DB 1; Length 957;
Best Local Similarity 32.3%; Pred. No. 2.2; Mismatches 23; Gaps 8;
Matches 42; Conservative 17; Indels 23; Gaps 8;

QY 13 GGGFAIPICGAMAIAGQINSG-----GGSPYHIGPTAFGLGV-VDNNGNGARVQVVG 66
DB 423 GGGFAGGIGAGGTPTGLNGGNGGASAVTGGNGGIGTGLNGGNGG--SGGIG 480
QY 67 SAPAASLGISTGDTVAVNG--APINSA---TAMDALNGHNGDVISTVWQKSGGTR 121
DB 481 AAKAGAGGVGS--GLLGLDGFNAPASTPLHTLQGVNL-----VYNEFPQTLTGRPL 532
QY 122 GNVYTLAEGPP 131
DB 533 GNG--ANGTP 540

RESULT 13
HMOA_HAEIN STANDARD; PRT; 466 AA.
AC P45129;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable periplasmic serine protease do/hnoA-like precursor
DE (EC 3.4.21.-).
GN H11259.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=727;
RN 11;
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7342800;
RA Fleischmann R.D., Adams M.D., White O., Claydon R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,

```

```

RA Scott J.D., Shirley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brindon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.,
RA "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: SEEMS TO BE A INTERMEDIATE FORMS BETWEEN E.COLI HTRA
CC (PROTEASE DO) AND HMOA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C. ALSO KNOWN AS THE
CC DEGP/DEGQ/DESS FAMILY.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U32805; AAC22906.1; -.
CC MEROPS: S01.274; -.
CC TIGR: H11259; -.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001940; Protease2C.
CC InterPro: IPR001254; Trypsln.
CC Pfam: PR00595; PDZ; 2.
CC DR Pfam: PR00089; trypsin; 1.
CC DR PRINTS: PR00834; PROTEASES2C.
CC SMART: SM00228; PDZ; 2.
CC DR PROSITE: PS50106; PDZ; 2.
CC KM Hydrolyase; Serine protease; Periplasmic; Signal; Complete proteome.
CC FT SIGNAL 1 29 POTENTIAL.
CC FT CHAIN 1 466 PROBABLE PERIPLASMIC SERINE PROTEASE
CC DO/HNOA-LIKE.
CC FT DOMAIN 270 361 PDZ 1.
CC FT DOMAIN 367 458 PDZ 2.
CC FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
CC FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (POTENTIAL).
CC SQ SEQUENCE 466 AA; 49434 MW; ED050A00047B551 CRC64;

Query Match 13.7%; Score 92.5; DB 1; Length 466;
Best Local Similarity 27.4%; Pred. No. 1.6;
Matches 32; Conservative 16; Mismatches 40; Indels 29; Gaps 3;

QY 13 GGGFAIPICGAMAIAGQINSG-----GGIRSGGSPYHIGPTAFGLGVVDNNGN 57
DB 253 GIAFAIPNSQASNIYQIILFEGGVRRG-----DLGKGGELNDLAKAFNVA 300
QY 58 --GARVQRVVGSAPASLSGISTGDTVAVNGAPINSATAMADALNGHHPDVISVTM 112
DB 301 QGGAGFVSEVLPKSAERAKAGDITIAMNGQKISSPAERAKIATTGAGKEISLTY 357

RESULT 14
DEGP_ARATH STANDARD; PRT; 448 AA.
ID DEGP_ARATH
AC Q9L010;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protease Do-like 8, chloroplast precursor (EC 3.4.21.-).
DE DEGP8 OR AT5G39830 OR K13H13.1.
GN Arabidopsis thaliana (Mouse-ear cress).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

```

NCBI_TaxID=3702;
[1]
SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.,
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[12]
SEQUENCE OF 91-113 AND 345-361.
RC STRAIN=CV, COLUMBIA;
RA Kieselbach T., Bystedt M., Schroeder W.P.;
RL Submitted (OCT-2000) to the SWISS-PROT data bank.
CC -!- FUNCTION: PROBABLE SERINE PROTEASE.
CC -!- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID LUMEN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
DEGP/DEGO/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AB024023; BAA98101.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; FALSE_NEG.
KW Hydroxylase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
FT TRANSIT 1 ?
FT CHAIN 1 90
FT DOMAIN 91 448
FT DOMAIN 152 333
FT DOMAIN 336 433
FT ACT_SITE 171 171
FT ACT_SITE 214 214
FT ACT_SITE 292 292
FT ACT_SITE 448 AA; 47492 MW; A986FC1387670AF CRC64;
SQ SEQUENCE 448 AA; 47492 MW; A986FC1387670AF CRC64;

Query Match 13.6%; Score 92; DB 1; Length 448;
Best Local Similarity 29.5%; Pred. No. 1.6; Mismatches 55; Indels 18; Gaps 4;
Matches 36; Conservative 13;
QY 10 SGGGGAIPICQAMALAGQIRSGGSPVHIGPTAFGLGVVDN---NGGARVORVVG 66
Db 316 TSAGVGFAPISSTVTKIYQLIQ--FSKVLKRGINIELAPDPYANOLNVRNGALVLPVG 373
QY 67 SAPASALG-----ISTGVDYTAVDGAPINSATMADALNGHHPGDVISVTWQTK 115
Db 374 KSLAKRAGIHPTSRGPNAGIVIGDIIVAVDDKPKKAKELMKILDEYSVD--KVLTKIK 431
QY 116 SG 117
Db 432 RG 433

RESULT 15
DEGP_BUCAI STANDARD; PRT; 478 AA.
AC P57323;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).

DEGP OR BU228.
GN Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium)
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
[1]
SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RT Buchnera sp. Aps.";
RL Nature 407:81-86(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
DEGP/DEGO/DEGS FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AP001118; BAB12943.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 1.
KW Hydroxylase; Serine protease; Signal; Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 478
FT DOMAIN 116 254
FT DOMAIN 281 372
FT DOMAIN 387 469
FT ACT_SITE 133 133
FT ACT_SITE 163 163
FT ACT_SITE 238 238
FT ACT_SITE 478 AA; 52230 MW; 868E8732CAC50629 CRC64;
SQ SEQUENCE 478 AA; 52230 MW; 868E8732CAC50629 CRC64;

Query Match 13.6%; Score 92; DB 1; Length 478;
Best Local Similarity 25.6%; Pred. No. 1.7; Mismatches 50; Indels 20; Gaps 4;
Matches 32; Conservative 23;
QY 13 GGGFAIPICQAMALAGQIRSGGSPVHIGPTAFGLGVVD-----NNGGARVORV 64
Db 265 GIGFAIPICQAMALAGQIRSGGSPVHIGPTAFGLGVVD-----NNGGARVORV 321
QY 65 VGSAPASALGISTGVDYTAVDGAPINSATMADALNGHHPGDVISVTWQTKSGGTRTG-- 122
Db 322 LPNSAFAEAGIKADIIISLNRKPISSFSRAEIGS-----LPVATKMEIGVFRGRI 375
QY 123 -NVTL 126
Db 376 KNITV 380

Search completed: August 6, 2002, 08:17:54
Job time: 164 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 08:12:10 ; Search time 16.84 Seconds
(without alignments)
191.460 Million cell updates/sec

Title: US-09-684-215A-4

Sequence: 1 TAAADNFQLSGGCGFAIP1.....QTKSGTRTGNVTLAEGPPA 132

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/prodata/2/1aa/PCBUS_COMB.pep.*
6: /cgn2_6/prodata/2/1aa/backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	670	99.3	132	US-08-818-112-66	Sequence 66, Appl
2	670	99.3	132	US-08-818-111-67	Sequence 67, Appl
3	670	99.3	132	US-09-056-556-66	Sequence 66, Appl
4	670	99.3	355	US-08-818-112-79	Sequence 79, Appl
5	670	99.3	355	US-08-818-111-80	Sequence 80, Appl
6	670	99.3	355	US-09-056-556-79	Sequence 79, Appl
7	200.5	29.7	580	US-08-818-112-75	Sequence 75, Appl
8	200.5	29.7	580	US-08-818-111-76	Sequence 76, Appl
9	200.5	29.7	580	US-09-056-556-75	Sequence 75, Appl
10	158.5	23.5	97	US-08-818-112-72	Sequence 72, Appl
11	158.5	23.5	97	US-08-818-111-73	Sequence 73, Appl
12	158.5	23.5	97	US-09-056-556-72	Sequence 72, Appl
13	148	21.9	30	US-09-439-313-484	Sequence 72, Appl
14	127.5	18.9	460	US-09-199-637A-132	Sequence 132, App
15	108	16.0	475	US-08-278-091-6	Sequence 6, Appl
16	108	16.0	475	US-08-483-859-6	Sequence 6, Appl
17	108	16.0	475	US-08-472-173-6	Sequence 6, Appl
18	108	16.0	475	US-08-487-167-6	Sequence 6, Appl
19	108	16.0	475	US-08-482-816-6	Sequence 6, Appl
20	108	16.0	475	US-08-296-149-6	Sequence 6, Appl
21	108	16.0	475	US-08-801-499-6	Sequence 6, Appl
22	108	16.0	475	US-08-615-271-6	Sequence 6, Appl
23	108	16.0	475	US-08-074-660-6	Sequence 6, Appl
24	108	16.0	475	US-09-074-660-6	Sequence 6, Appl
25	108	16.0	475	US-09-106-468-6	Sequence 6, Appl
26	108	16.0	475	US-09-106-468A-6	Sequence 6, Appl
27	108	16.0	475	US-09-106-467-6	Sequence 6, Appl

28	105	15.6	475	1	US-08-350-741-2	Sequence 2, Appl
29	105	15.6	475	2	US-08-463-875A-2	Sequence 2, Appl
30	102.5	15.2	472	1	US-08-278-091-5	Sequence 5, Appl
31	102.5	15.2	472	1	US-08-483-859-5	Sequence 5, Appl
32	102.5	15.2	472	1	US-08-472-173-5	Sequence 5, Appl
33	102.5	15.2	472	2	US-08-487-167-5	Sequence 5, Appl
34	102.5	15.2	472	2	US-08-482-816-5	Sequence 5, Appl
35	102.5	15.2	472	2	US-08-296-149-5	Sequence 5, Appl
36	102.5	15.2	472	2	US-08-801-499-5	Sequence 5, Appl
37	102.5	15.2	472	2	US-08-615-271-5	Sequence 5, Appl
38	102.5	15.2	472	3	US-09-074-660-5	Sequence 5, Appl
39	102.5	15.2	472	3	US-09-074-659-5	Sequence 5, Appl
40	102.5	15.2	472	3	US-09-106-468-5	Sequence 5, Appl
41	102.5	15.2	472	4	US-09-106-468A-5	Sequence 5, Appl
42	102.5	15.2	472	4	US-09-106-467-5	Sequence 5, Appl
43	102.5	15.2	472	1	US-08-923-454A-14	Sequence 14, Appl
44	102.5	15.2	492	1	US-08-350-741-3	Sequence 3, Appl
45	102.5	15.2	492	2	US-08-463-875A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1
US-08-818-112-66
Sequence 66, Application US/08818112
Patent No. 6290369
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedicik, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESSES:
ADDRESS: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-66

Query Match 99.3%; Score 670; DB 4; Length 132;
Best Local Similarity 99.2%; Pred. No. 2.4e-62;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-79

Query Match 99.3%; Score 670; DB 4; Length 355;
Best Local Similarity 99.2%; Pred. No. 8.9e-62;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGGFAIPGQAMAIAGQIRSGGSPYHIGPFAFLGLGVNDNNGAR 60
|||||
DB 224 TAASDNFOLSGGGGFAIPGQAMAIAGQIRSGGSPYHIGPFAFLGLGVNDNNGAR 283

QY 61 VORVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWOTKSGCTR 120
|||||
DB 284 VORVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWOTKSGCTR 343

QY 121 TGNVTLAEGPPA 132
|||||
DB 344 TGNVTLAEGPPA 355

RESULT 5
US-08-818-111-80
Sequence 80, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 80:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-80

Query Match 99.3%; Score 670; DB 4; Length 355;
Best Local Similarity 99.2%; Pred. No. 8.9e-62;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGGFAIPGQAMAIAGQIRSGGSPYHIGPFAFLGLGVNDNNGAR 60
|||||
DB 224 TAASDNFOLSGGGGFAIPGQAMAIAGQIRSGGSPYHIGPFAFLGLGVNDNNGAR 283

QY 61 VORVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWOTKSGCTR 120
|||||
DB 284 VORVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWOTKSGCTR 343

QY 121 TGNVTLAEGPPA 132
|||||
DB 344 TGNVTLAEGPPA 355

RESULT 6
US-09-056-556-79
Sequence 79, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392

```
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-036-536-79

Query Match
Best Local Similarity 99.3%; Score 670; DB 4; Length 355;
Matches 131; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGNGAR 60
DB 224 TAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGNGAR 283
QY 61 VQRVGSAFASISGISTGDIYTAVDGAPINSATMADALNGHHPDVISYTWQTSKGR 120
DB 284 VQRVGSAFASISGISTGDIYTAVDGAPINSATMADALNGHHPDVISYTWQTSKGR 343
QY 121 TGNVTLAEGPPA 132
DB 344 TGNVTLAEGPPA 355

RESULT 7
US-08-818-112-75
Sequence 75, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 580 amino acids
TYPE: amino acid
STRANDEDNESS: single
```

```
TOPOLOGY: linear
US-08-818-112-75

Query Match
Best Local Similarity 29.7%; Score 200.5; DB 4; Length 580;
Matches 51; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 1 TAASDNFQLSQG--GGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVV--DNNGN 57
DB 453 TLGADSDAOSGSIGLGFALPVDARKRIADELISTGA-----SHASLGVOYTNDRKDP 506
QY 58 GARVQRVGSAFASISGISTGDIYTAVDGAPINSATMADALNGHHPDVISYTWQTSKSG 117
DB 507 GAKIVIVAGGAANAANQVPRGVYTVKDDRPINSADALVAAYVSKAPGATVALTFDDPSG 566
QY 118 GTRTGNVTLAE 128
DB 567 GSRTVQVTLGK 577

RESULT 8
US-08-818-111-76
Sequence 76, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 76:
SEQUENCE CHARACTERISTICS:
LENGTH: 580 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-76

Query Match
Best Local Similarity 29.7%; Score 200.5; DB 4; Length 580;
Matches 51; Conservative 21; Mismatches 50; Indels 9; Gaps 3;

QY 1 TAASDNFQLSQG--GGGFAIPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVV--DNNGN 57
```

D b	453	TIGADSDAQSGSIGLEFALIPVDQAKRIADELLISTGCA-----SHAHSLSGVQYTNKDTP	506
Q y	58	GARQRVRVGCNAPASLGLSTGYDVTANDGAPINSATMTDALNGHHEGDVYSTWTQSKG	117
D b	507	GAKEVEYVAGGAANAGVPKGVVYTKVDRPINSADALVAAYNSKAPGATVALTFDPDSG	566
Q y	118	GTRTGNVTLAEL	128
D b	567	GSRTGVQYTLGK	577

RESULT 9
US-09-056-556-75

```

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/056,556
FILING DATE: 07-APR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MakI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.457
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 580 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IS-09-056-556-75

```

Query Match	29.7%	Score 200.5	DB 4	Length 580
Best Local Similarity	38.9%	Pred. No. 7,7e-13		
Matches 51	Conservative 21	Mismatches 50	Indels 9	Gaps 3

QY 118 CTRTGNTLAE 128
1:11 111:
Db 567 GSRTVQYTLGK 577
507 GAKTVEYVAGGAANAGVPRKGVYTKVDDPRINSADALVAAVESKAKGATVALLTEQPSG 5666

RESULT 10
US-08-818-112-72
; Sequence 72, Application US/08818112

Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skelky, Yasin A.W.
APPLICANT: Dillon, David C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedrick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
City: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ. ID NO: 72:
SEQUENCE CHARACTERISTICS:
LENGTH: 97 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

Query Match	23.58;	Score 158.5;	DB 4;	Length 97;
Best Local Similarity	38.78;	Pred. No. 1.7e-09;		
Matches	36;	Conservative 16;	Mismatches 38;	Indels 3; Gaps 1

```
QY      96 ADALNGHHPGDVISVTWQTKSGGTRTGNVTLAE 128  
       :  ||::||| |||:  
Db      62 VAAVRSKAPGATVALTFQDPSSGSRIVQYLIGK 94
```

RESULT 11
 US-08-818-111-73
 ; Sequence 73, Application US/08818111
 ; Patent No. 6338852
 ;
 ; GENERAL INFORMATION:
 ;
 APPLICANT: Reed, Steven G.
 APPLICANT: Skeiky, Yasir A.W.
 APPLICANT: Dillon, David C.
 APPLICANT: Campos-Neto, Antonia
 APPLICANT: Houghton, Raymond
 APPLICANT: Vedicik, Thomas S.
 APPLICANT: Twardzik, Daniel R.
 TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 TUBERCULOSIS
 NUMBER OF SEQUENCES: 148
 ;
 ; CORRESPONDENCE ADDRESS:
 ;
 ADDRESSEE: SEED and BERRY LLP
 ;
 ;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 6, 2002, 08:14:20 ; Search time 26.07 Seconds
(without alignments)
875.924 Million cell updates/sec

Title: US-09-684-215A-4

Perfect score: 675
Sequence: 1 TAAADNFQSLQSGGQGFAPL.....QTKSGGTRIGNVTLAGCPA 132

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SPTREMBL_19:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being predicted, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	675	100.0	355	16 0071175	0071175 mycobacteri
2	474.5	70.3	361	2 050320	050320 mycobacteri
3	459.5	68.1	354	16 09C9CY9	09C9CY9 mycobacteri
4	199.5	29.6	464	16 053896	053896 mycobacteri
5	188.5	27.9	382	16 09CDE7	09CDE7 mycobacteri
6	188.5	27.9	452	2 0925G6	0925G6 mycobacteri
7	161.5	23.9	542	2 09FBR9	09FBR9 streptomyc
8	154	22.8	519	2 093J30	093J30 streptomyc
9	139	20.6	394	16 072780	072780 streptomyc
10	138.5	20.5	362	2 09S2R5	09S2R5 streptomyc
11	131	19.4	441	16 09RTK4	09RTK4 streptomyc
12	128.5	19.0	514	16 09PBA3	09PBA3 xylella fas
13	127.5	18.9	371	2 031388	031388 bradyrhizob
14	127.5	18.9	474	2 09ALST	09ALST pseudomonas
15	122.5	18.1	474	16 057155	057155 pseudomonas
16	121	17.9	513	16 098CS8	098CS8 rhizobium 1

17	120.5	17.9	465	16 0920E6	0920E6 rhizobium m
18	119.5	17.7	203	2 09X840	09X840 streptomyc
19	118	17.5	516	16 0985F9	0985F9 rhizobium 1
20	117.5	17.4	530	16 09A4S2	09A4S2 caulobacter
21	117	17.3	473	2 044476	044476 azotobacter
22	113	16.7	393	16 097N37	097N37 streptococ
23	113	16.7	397	2 006670	006670 streptococ
24	112	16.6	500	2 044652	044652 bruceella ad
25	111	16.4	481	2 09AQD1	09AQD1 pseudomonas
26	111	16.4	499	16 09JVT1	09JVT1 neisseria m
27	108	16.0	348	16 097LU1	097LU1 clostridium
28	108	16.0	472	16 09A8R9	09A8R9 caulobacter
29	107.5	15.9	452	16 P73354	P73354 streptocyst
30	104.5	15.5	481	2 09LRK0	09LRK0 shigella so
31	104	15.4	397	2 093F87	093F87 streptococ
32	103.5	15.3	407	16 095XG9	095XG9 streptococ
33	102.5	15.2	428	16 098N31	098N31 rhizobium 1
34	102	15.1	500	2 09KJN6	09KJN6 myxococcus
35	100.5	14.9	453	16 067436	067436 aquifex ao
36	100.5	14.9	459	16 09WZ41	09WZ41 thermotoga
37	100	14.8	398	16 083752	083752 treponema p
38	100	14.8	652	16 092G62	092G62 rickettsia
39	99.5	14.7	224	16 035021	035021 bacillus su
40	99.5	14.7	458	2 09R911	09R911 bacillus su
41	98.5	14.6	481	16 09PGL3	09PGL3 xylella fas
42	98	14.5	476	16 09ZM18	09ZM18 helicobacte
43	97.5	14.4	453	2 09FDP1	09FDP1 aeromonas h
44	95.5	14.1	290	5 021294	021294 caenorhabdi
45	95.5	14.1	404	16 033351	033351 mycobacteri

ALIGNMENTS

RESULT 1	007175	PRELIMINARY;	PRT;	355 AA.
ID	007175			
AC	007175			
DT	01-JUL-1997 (TREMBLrel. 04, Created)			
DT	01-JUL-1997 (TREMBLrel. 04, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	HYPOTHETICAL 34.9 KDA PROTEIN.			
GN	PEPA OR RV0125 OR MTC1418B.07.			
OS	Mycobacterium tuberculosis.			
CC	Bacteria: Firmicutes; Actinobacteria; Actinobacteridae;			
CC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-H37RV;			
RX	MEBLINE-98295987; PubMed-9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltham R., Gentles S., Hamlin N., Holtroyd S.,			
RA	Hornsey T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutten J.E., Seeger K., Skelton S., Squares R.,			
RA	Suston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
CC	-1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.			
DR	EMBL: Z96071; CAB09453.1; -			
DR	MEROPS: S01.0PC; -			
DR	TubercuList: RV0125; -			
DR	InterPro: IPR001478; PDZ.			
DR	InterPro: IPR001254; Trypsin.			
DR	Pfam: PF00595; PDZ; 1.			
DR	Pfam: PF00088; trypsin; 1.			
DR	SMART: SM00228; PDZ; 1.			
DR	PROSITE: PS50106; PDZ; 1.			
DR	PROSITE: PS50240; TRYPSIN_DOM; 1.			

DR PROSITE: PS00135; TRYPSIN_SER; UNKNOWN.1.
 KM Complete proteome: Hydrolyase; Hypothetical protein: Serine protease.
 SQ SEQUENCE 355 AA; 34926 MW; 16CB9E21A97BF192 CRC64;

Query Match 100.0%; Score 675; DB 16; Length 355;
 Best Local Similarity 100.0%; Pred. No. 1.5e-41;
 Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TAASDNFOLSGGGGFAIPGQAMATAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
 DB 224 TAASDNFOLSGGGGFAIPGQAMATAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 283
 QY 61 VQRVGSAAPAASIGISTGDIYAVDGPAPINSATAMADALNGHHPGDVISTVWQTKSGGTR 120
 DB 284 VQRVGSAAPAASIGISTGDIYAVDGPAPINSATAMADALNGHHPGDVISTVWQTKSGGTR 343
 QY 121 TGNVTLAEGPPA 132
 DB 344 TGNVTLAEGPPA 355

RESULT 2
 Q50320 PRELIMINARY; PRT; 361 AA.
 AC 050320;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 34KDA PROTEIN PRECURSOR.
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_Taxid=1770;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=JD88/107;
 RX MEDLINE=95003449; PubMed=7921248;
 RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
 RT "Identification and characterization of a putative serine protease
 expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
 RL Microbiology 140:1977-1982(1994).
 DR EMBL: 223092; CAA80638.1; -;
 DR MEROPS: S01.UPC; -;
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ.1.
 DR Pfam: PF00089; Trypsin.1.
 DR SMART: SM00228; PDZ.1.
 DR PROSITE: PS50106; PDZ.1.
 DR PROSITE: PS50240; TRYPSIN_DOM.1.
 DR Hydrolyase; Serine protease; Signal.
 KW SIGNAL.1
 FT SIGNAL.1
 SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FDF3C411 CRC64;

Query Match 70.3%; Score 474.5; DB 2; Length 361;
 Best Local Similarity 69.7%; Pred. No. 4.9e-27;
 Matches 92; Conservative 17; Mismatches 22; Indels 1; Gaps 1;

QY 1 TAASDNFOLSGGGGFAIPGQAMATAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
 DB 231 TAATDSTKMS-GGQGFALPFGQAMEVGAIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 289
 QY 61 VQRVGSAAPAASIGISTGDIYAVDGPAPINSATAMADALNGHHPGDVISTVWQTKSGGTR 120
 DB 290 VQRVGSAAPAASIGISTGDIYAVDGPAPINSATAMADALNGHHPGDVISTVWQTKSGGTR 349
 QY 121 TGNVTLAEGPPA 132
 DB 350 TGNVTLAEGPPA 361

RESULT 3
 Q9CCY9 PRELIMINARY; PRT; 354 AA.
 AC Q9CCY9;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PROBABLE SECRETED SERINE PROTEASE.
 GN M2659.
 OS Mycobacterium leprae.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 NCBI_Taxid=1769;
 [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=TN.
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eigemeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajadream M., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;

RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 409:1007-1011(2001).
 CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AL583926; CAC32191.1; -;
 DR MEROPS: S01.UPC; -;
 DR Lepronia; M2659; -;
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR000126; Ser.proteas_v8.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00595; PDZ.1.
 DR Pfam: PF00089; Trypsin.1.
 DR PRINTS: PRO0834; PROTEASES2C.
 DR PRINTS: PRO0839; V8PROTEASE.
 DR SMART: SM00228; PDZ.1.
 DR PROSITE: PS50106; PDZ.1.
 DR PROSITE: PS50240; TRYPSIN_DOM.1.
 DR Complete proteome: Hydrolyase; Protease; Serine protease.
 KW SEQUENCE 354 AA; 35265 MW; 612F23261BC9E4A4 CRC64;

Query Match 68.1%; Score 459.5; DB 16; Length 354;
 Best Local Similarity 66.7%; Pred. No. 5.9e-26;
 Matches 88; Conservative 17; Mismatches 26; Indels 1; Gaps 1;

QY 1 TAASDNFOLSGGGGFAIPGQAMATAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 60
 DB 224 TAATDNTKM-LGGQGFALPFGQAMEVGAIRSGGSPVTHIGPTAFGLGVVDNNGNGAR 282
 QY 61 VQRVGSAAPAASIGISTGDIYAVDGPAPINSATAMADALNGHHPGDVISTVWQTKSGGTR 120
 DB 283 VARVATGPAAMAGISVDITSDVGPISBATMTNVLVPHHGGTAVAVNRSAGGDL 342
 QY 121 TGNVTLAEGPPA 132
 DB 343 TGNVTLAEGPPA 354
 RESULT 4
 O53896 PRELIMINARY; PRT; 464 AA.
 AC O53896;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE PUTATIVE SERINE PROTEASE.
 GN RV0983 OR MTV044.11.

QW Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinomycetales;
OC Actinomycetales; Corynebacteriineae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA Badcock K., Basham D., Blyth D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feldwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jajels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Salston J.E., Taylor K., Whitehead S., Barrrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: AL021999; CAI17582.1; -.
DR MEROPS: S01.UPC; -.
DR Tuberculist: RV0983; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF000595; PDZ. 1.
DR Pfam: PF000895; trypsin. 1.
DR SMART: SM00228; PDZ. 1.
DR PROSITE: PSS0106; PDZ. 1.
DR PROSITE: PSS0240; TRYPsin_DOM. 1.
DR Complete proteome; Hydrolase; Serine protease.
QW SEQUENCE 464 AA; 46452 MW; AE93BFFCC53E1EC8F CRC64;

Query Match	29.6%;	Score 199.5;	DB 16;	Length 464;
Best Local Similarity	38.9%;	Pred. No. 5.4e-07;		
Matches 51; Conservative	21;	Mismatches 50;	Indels 9;	Gaps 3

[illegible]

	RESULT	5
09CD67		
ID	09CD67	PRELIMINARY; PRT; 382 AA.
AC	09CD67;	
DT	01-JUN-2001 (TREMBLrel. 17, Created)	
DT	01-JUN-2001 (TREMBLrel. 17, last sequence update)	
DT	01-DEC-2001 (TREMBLrel. 19, last annotation update)	
DE	POSSIBLE SECRETED SERINE PROTEASE.	
GN	MT0176.	
OS	Mycobacterium leprae.	
OC	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;	
CC	Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID:1769;	
RA	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-TN:	
RX	MEDLINE-21128732; Pubmed-11234002;	
RA	Cole S.T., Elgmeyer K., Parkhill J., James K.D., Thomson N.R.,	
RA	Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,	
RA	Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA	Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,	
RA	Holtrov� S., Hornsby T., Jagers K., Lacroix C., Maclean J., Moule S.,	

RA Rutherford K.M.,
RA Rutter S., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Smith S., Seeger K., Simón S., Simmonds M., Skellton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrett B.G.,
RT "Massive gene decay in the leprosy bacillus.",
RL Nature 409:1007-1011(2001).
CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583917; CAC29684.1; -.
DR MEROPS; S01.UPC; -.
DR Leproma; ML0176; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00069; trypsin; 1.
DR PRINTS; PR00834; PROTEASE2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
DR Complete proteome; HydroLase; Protease; Serine protease.
SQ Sequence 382 AA; 37084 MW; 3DD8DDDAE32A80D CRC64;

Query Match	27.9%	Score 188.5;	DB 16;	Length 382;
Best Local Similarity	39.8%	Pred. No. 2.7e-06;		
Matches 51; Conservative	19;	Mismatches 51;	Indels 7;	Gaps 3;

```

Oy  2 AASDNFQLSGGCGCPATPIGAMMAAQIBSGGSGSPYHIGPTFLGLGVDDNCG--GAR 60
    250 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db  258 ADGSGAGGSGSGIGLGVAPVDQAKRRAEDLLISTG--KATH---ASLGVQATDKGTGTGAK 311
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Oy  61 VQRYVGSAPASLGISTDVITVADGAFINSATMADALNGHFGDVIYVTWTKSGGTR 120
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
Db  312 VMDVYAGGAANAANVAPKCVLTLTKDDRLIISSDADLLVAARSKAGDVSLSLYDQDSGSSR 379
    1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1

```

QY	121	TGNVTLAE	128
		:	
Db	372	TVQVTLGK	379

```

RESULT      6
092556      PRELIMINARY;      PRT;      452 AA.
ID      092556;
AC      092556;
DT      01-MAY-1999 (TREMBLrel. 10, Created)
DT      01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      PUTATIVE SERINE PROTEASE.
GN      MCB373.28.
OS      Mycobacterium leprae.
OC      Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC      Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX      NCBI_TaxID=1769;
RN      11]
RP      SEQUENCE FROM N.A.
RA      Harris D., Taylor K.;
RL      Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN      12]
RP      SEQUENCE FROM N.A.
RA      James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL      Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN      13]
RP      SEQUENCE FROM N.A.
RA      MEDLIN=93188700; PubMed=8446027;
RA      Elgmæier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT      "Use of an ordered cosmid library to deduce the genomic organization
RT      of Mycobacterium leprae.";
RL      Mol. Microbiol. 7:197-206(1993).
DR      EMBL: AL035500; CAB3690.1.-.
DR      MEROPS: S01.UPC.-.
DR      InterPro: IPR001478; PDZ.
DR      InterPro: IPR001940; Protease2C.
DR      InterPro: IPR001254; Trypsin.
DR      Pfam: PF00595; PDZ; 1.

```



```

P72780
ID P72780 PRELIMINARY; PRT: 394 AA.
AC P72780;
DT 01-FEB-1997 (Tremblrel. 02, Created)
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE PROTEASE HHOA.
GN HHOA OR SL1679.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Sugitani M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shilpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -I- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: D90900; BAA16795.1;
DR MEROPS: S01.274;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ. 1.
DR Pfam: PF00089; trypsin. 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ. 1.
DR PROSITE: PS50106; PDZ. 1.
DR PROSITE: PS50240; TRYPSIN_DOM. 1.
DR Complete proteome: Hydrolase; Protease; Serine protease.
KW SEQUENCE 394 AA; 41336 MW; AD59D94811B8F57B CRC64;
SQ

Query Match 20.6%; Score 139; DB 16; Length 394;
Best Local Similarity 34.6%; Pred. No. 0.011;
Matches 45; Conservative 10; Mismatches 43; Indels 32; Gaps 5;

OY 13 GGGFAIPIGQAMAIAGQIRSGGSPYVHIGPTAFLGLGV---VDNNGN-----G 58
DB 261 GIGFAIPIDAKAIONLACGIVPHRYIG-VQMMNITVDQAQNNRNPNSEFFIPEVDG 319
OY 59 ARVQRVVGSAPASLIGSTGDIVITAVDGAIPINSATAM-----ADALNGHHP 104
DB 320 ILVMRYLPGTPAERAGJRRGDVIVAVDGTPIISDGARLQRIVEQAGLNKALKLDLGRDR 379
OY 105 GDVISTVWQT 114
DB 380 ---LSLTVQT 386

RESULT 10
O9S2R5 PRELIMINARY; PRT: 362 AA.
AC O9S2R5;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE SECRETED PROTEASE.
GN SC5F7.30.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-A3(2);
RA Seeger K., Harris D.;

```

```

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Bentley S.D., Parkhill J., Barrett B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kleser H.M., Denapite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL096872; CAB51255.1;
DR MEROPS: S01.UPC;
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ. 1.
DR Pfam: PF00089; trypsin. 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ. 1.
DR PROSITE: PS50106; PDZ. 1.
DR PROSITE: PS50240; TRYPSIN_DOM. 1.
DR Hydrolase; Protease; Serine protease.
KW SEQUENCE 362 AA; 36381 MW; 71CC41F943A6D03B CRC64;
SQ

Query Match 20.5%; Score 138.5; DB 2; Length 362;
Best Local Similarity 35.5%; Pred. No. 0.011;
Matches 43; Conservative 14; Mismatches 55; Indels 9; Gaps 4;

OY 13 GGGFAIPIGQAMAIAGQIRSGG---GSPYVHIGPTAFLGLGVNDNNGN--GARVQRVVGS 67
DB 245 GIGFAIPASVTVYAGQIVADGKVTDSGRALGITA---RTVVDSTYPARAAVVEVSDG 301
OY 68 APAASLGISTGDIVITAVDGAIPINSATAMADALNGHHPGDIVISVWQTKSGGTRGNVTLA 127
DB 302 GAADDGRLRGDVLVVKLGDDITITSLSEALASMRGDRFRKYV-TRDGKEHRAEVTLG 360
OY 128 E 128
DB 361 E 361

RESULT 11
O9RTK4 PRELIMINARY; PRT: 441 AA.
AC O9RTK4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PERIPLASMIC SERINE PROTEASE DO, PUTATIVE.
GN DR1756.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-RL;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eissen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RL.";
RL Science 286:1571-1577(1999).

```


	Matches	33;	Conservative	24;	Mismatches	55;	Indels	7;	Gaps	4;
Qy	13	GCGFAIRPIGQAAMAIAGTIRSGGSPYTHIGETPAFLGLGV--VDNNGGARVQRVVGSA	P69							
Dd	243	GIGFAIVPMARRMEQILOYGEVRRCIG-ISRIDLGLDLAKESYOGALIAETIASGP	301							
Qy	70	AASLGISTGVITVAVDGAPINSATAMADLGNHHPGVISTWTQTKSGCTGTGNTTLA	129							
Dd	302	AEQAGLGKGDIVKAVDGPTRKSASQLRNLIGLPVSGRVELRFE-RNGAASVASVEY--G	358							
Qy	130	p 130								
Dd	359	p 359								
RESULT	14									
ID	09ALSL1	PRELIMINARY;	PRT;	474	AA.					
AC	09ALSL1									
DT	01-JUN-2001	(TREMBLrel. 17, Created)								
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)								
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)								
DE	MUCD.									
GN	MUCD.									
OS	Pseudomonas aeruginosa.									
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;									
CC	Pseudomonas.									
OX	NCBI_TaxID=287;									
RN	111									
RP	SEQUENCE FROM N.A.									
RC	STRAIN-UICBP-PAL4;									
RA	Torrey P.S., Ralme L.G., Tan M., Ausubel F.M.;									
RT	"The Roles of mucD and Alginate in the Virulence of Pseudomonas									
FT	aeruginosa In Plants, Nematodes, and Mice.";									
RL	Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.									
DR	EMBL: AF343973; AAK11276.1; -									
DR	InterPro: IPR001478; PDZ.									
DR	InterPro: IPR001940; Proteinase2C.									
DR	InterPro: IPR001254; Trypsin.									
DR	Pfam: PF00595; PDZ_2.									
DR	Pfam: PF00089; trypsin_1.									
DR	PRINTS: PR00834; PROTEASES2C.									
DR	SMART: SM00288; PDZ_2.									
DR	PROSITE: PS50106; PDZ_2.									
DR	PROSITE: PS50240; TRYPsin_DOM; 1.									
KW	Hydrolase; Serine protease.									
SO	SEQUENCE 474 AA; 50349 MW; A2ZFD4338B859D4C CRC64;									
Query Match	18.9%;	Score 127.5;	DB 2;	Length 474;						
Best Local Similarity	31.2%;	Pred. No. 0.09;								
Matches	43;	Conservative	21;	Mismatches	51;	Indels	23;	Gaps	5;	
Qy	4	SDNFLDSG--GCGFAIRPIGQAAMAIAGTIRSGGSPYTHIGETPAFLGLGVVDNN-----	55							
Dd	233	SQIFRRSGGFMLSPALPDIVALNVADDLKRG--KVSRG--WLGVVIQEVNKDLAES	286							
Qy	56	-----GNGARVQRVVGSPAASLSIGTGVTVAVDGAPINSATAMADLGNHHPGVISY	110							
Dd	287	FGLDKPSGALVAQLVEDPGRAKGGLQVGYDLSLWGQSINESADLRPHLVGMKKPGDKINL	346							
Qy	111	----TWQTKSGGTRTGNV	124							
Dd	347	DVIRNGGRKRSLSMAVGNL	364							
RESULT	15									
ID	057155	PRELIMINARY;	PRT;	474	AA.					
AC	057155									
DT	01-NOV-1996	(TREMBLrel. 01, Created)								
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)								
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)								

```

DE MUCD (SERINE PROTEASE MUCD).
OS MUCD OR PA0766.
OC Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PAOI;
RX MEDLINE=93391355; PubMed=8378309;
RA Martin D.W., Schurr M.J., Mudd M.H., Govan J.R., Holloway B.W.,
RA Deretic V.;
RT "Mechanism of conversion to mucoidy in Pseudomonas aeruginosa
RT infecting cystic fibrosis patients.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:8377-8381(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PAOI;
RX MEDLINE=95286510; PubMed=7768826;
RA Yu H., Schurr M.J., Deretic V.;
RT "Functional equivalence of Escherichia coli sigma E and Pseudomonas
RT aeruginosa AlgU: E. coli rpoE restores mucoidy and reduces sensitivity
RT to reactive oxygen intermediates in algU mutants of P. aeruginosa.";
RL J. Bacteriol. 177:3259-3268(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PAOI;
RX MEDLINE=96134987; PubMed=8550474;
RA Boucher J.C., Martinez-Salazar J., Schurr M.J., Mudd M.H., Yu H.,
RA Deretic V.;
RT "Two distinct loci affecting conversion to mucoidy in Pseudomonas
RT aeruginosa in cystic fibrosis encode homologs of the serine protease
RT HtrA.";
RL J. Bacteriol. 178:511-523(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey R.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
RA Garner R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardén K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saler M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -I - SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: U49151; AAC43718.1; -.
DR EMBL: U32853; AAC43676.1; -.
DR EMBL: AE004511; AAC04155.1; -.
DR MEROPS: S01.UPC; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00595; PDZ; 2.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Complete proteome; Hydrolase; Protease; Serine protease.
SO SEQUENCE 474 AA; 50321 MW; 8AB5D2A89867BEBE CRC64;

```

```

OY 56 -----GNGARVORVGSAPASLIGISTGDVITAVDGA PINSATAMADALNGHHPGDVISV 110
      :||| |::| ||| |::| ||| |::| ||| |::| ||| |::| ||| |::|
Db 287 FGLDKPSGALVAQLVEBDGPAKGGLOVGDVITLSLNGOSINESADLPHLVGNMKGPGDKINTL 346
OY 111 ----TWOTKSGGTRTGNV 124
      | || |::|
Db 347 DVIRNGORRKSLSMAVSSL 364

```

Search completed: August 6, 2002, 08:17:32
 Job time: 192 sec